THEORY OF LIGHT

MACMILLAN AND CO., Limited
LONDON · BOMBAY · CALCUTTA · MADRAS
MELBOURNE

THE MACMILLAN COMPANY
NEW YORK · BOSTON · CHICAGO
DALLAS · ATLANTA · SAN FRANCISCO

THE MACMILLAN COMPANY
OF CANADA, LIMITED
TORONTO

THEORY OF LIGHT

Being Volume IV of
"INTRODUCTION TO THEORETICAL PHYSICS"

BY

MAX PLANCK

Nobel Laureate, Foreign Member Royal Society, Professor of Theoretical Physics, University of Berlin, and President of the Kaiser Wilhelm Research Institute.

TRANSLATED BY

HENRY L. BROSE,
M.A., D.Phil. (Oxon.), D.Sc.

Lancashire-Spencer Professor of Physics,
University College, Nottingham.

MACMILLAN AND CO., LIMITED,
ST. MARTIN'S STREET, LONDON
1932

PRINTED IN GREAT BRITAIN.

PREFACE TO THE FIRST GERMAN EDITION

THE scheme of arrangement of the present volume runs along precisely the same lines as those which guided me in preparing the three first volumes of the present series, which aims at giving a thorough introduction to theoretical physics. In view of the great area over which theoretical optics now extends, we have in this case, too, been able to deal with only a meagre selection from the very abundant material that is available. The choice made was governed first and foremost by the desire to restrict the discussion to the framework of the classical theory as applied to bodies of continuous space-distribution. This enabled me to lay greater stress on the system used in arranging and developing the theorems, and on their links with the other branches of theoretical physics. For this reason there are numerous references to the preceding volumes of this series; the Roman figure I refers to the volume on General Mechanics, II to that on the Mechanics of Deformable Bodies, III to that on the Theory of Electricity and Magnetism.

Although the assumption of matter having absolutely continuous properties could be maintained in all the preceding volumes, it was found necessary here to overstep this assumption in treating the problem of dispersion. And since it is impossible to leave out dispersion in giving an account of theoretical optics, I have included a first introduction to the atomic point of view in the last part of this volume and have seized this opportunity to make an attempt to describe the way in which it links up naturally with quantum mechanics. For the circumstance that access to quantum mechanics, as well as to

the theory of relativity, can be gained most readily from the side of classical theory by making an appropriate generalization seems to be indicated not only from the didactic point of view but also from considerations of concreteness of expression.

An index to the definitions and theorems is appended.

MAX PLANCK.

Berlin-Grunewald,
July 1927.

PREFACE TO THE SECOND EDITION

THE new edition differs from the old only in a few minor points and in having a few additions.

MAX PLANCK.

Berlin-Grunewald,
December 1930.

CONTENTS

INTRODUCTION

§ 1. PHYSICAL optics is a special department of electrodynamics—namely, that which comprises the laws of rapidly varying fields. Its particular significance consists in the fact that it represents the branch of physics in which the most refined measurements are possible, and which consequently enables us to penetrate furthest into the details of physical phenomena. At the same time, optics presents a clearer illustration than any other branch of physics of the peculiar tendency of progressive scientific research to leave the original point of departure—namely, the specific sense-impressions—and to place physical concepts on more objective foundations. For, whereas the most important optical concepts, those of light and colour, were originally derived from the impressions on our eyes, these concepts have nothing at all to do with the immediate sensation of sight in present-day physics, but relate rather to electromagnetic waves and vibration periods—a trend of development which has justified itself in the abundant fruit which it has borne.

§ 2. We can progress most easily by linking up with the general system of Maxwell's equations for the electromagnetic field in stationary bodies, particularly if we use the special form which they assume for transparent and non-magnetic bodies. Since the transparency of a body is associated with the condition that no transformation of electromagnetic energy into heat occurs in it, all transparent bodies are electrical insulators in which the vector $\boldsymbol{J}$ of the electric flux vanishes everywhere and at all times. Besides excluding conductors, this also excludes strongly magnetizable bodies; for other bodies we may, without introducing an appreciable error, identify the

magnetic induction $\boldsymbol{B}$ with the magnetic intensity of field $\boldsymbol{H}$. Then, by III (31) the field equations assume the simple form :

$$\dot{\boldsymbol{D}} = c \operatorname{curl} \boldsymbol{H}, \quad \dot{\boldsymbol{H}} = -c \operatorname{curl} \boldsymbol{E} \quad . \quad . \quad . \quad (1)$$

together with the supplementary equations III (49) and (51) :

$$\operatorname{div} \boldsymbol{D} = 0, \quad \operatorname{div} \boldsymbol{H} = 0 \quad . \quad . \quad . \quad . \quad (2)$$

Here $\boldsymbol{E}$ denotes the electric intensity of field, $\boldsymbol{H}$ the magnetic intensity of field, $\boldsymbol{D}$ the electric induction, c the critical velocity, all quantities being measured in the so-called Gaussian system of units (III, § 7).

The above system of equations embraces the optics of all transparent substances. But the variables that occur in them play the part only of auxiliary quantities, since they are not directly measured. There is one quantity, to determine which is the goal of all optical measurements and to calculate which is therefore the proper task of every optical theory. This quantity is the vector of the electromagnetic flux of energy :

$$\boldsymbol{S} = \frac{c}{4\pi}[\boldsymbol{E}, \boldsymbol{H}] \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

which gives the intensity and direction of the intensity of radiation [see III (26)].

For the subsequent treatment of these equations we have to take into consideration the particular relation which connects the vector of the electric intensity of field $\boldsymbol{E}$ with the vector of electric induction $\boldsymbol{D}$ and which endows a substance with its characteristic optical behaviour. Accordingly we find it appropriate to divide the material into three parts, so that we successively discuss the optics of isotropic homogeneous bodies, the optics of crystals and the optics of non-homogeneous bodies in which the phenomena of dispersion and absorption are included.

PART ONE

OPTICS OF ISOTROPIC AND HOMOGENEOUS BODIES

CHAPTER I

REFLECTION AND REFRACTION

§ 3. In the case of an isotropic and homogeneous substance the relation between electric induction and electric intensity of field is expressed by the equation III (28):

$$\boldsymbol{D} = \epsilon \, . \, \boldsymbol{E} \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where ϵ denotes the dielectric constant. The field-equations (1) then become:

$$\epsilon\dot{\boldsymbol{E}} = c \operatorname{curl} \boldsymbol{H}, \ \dot{\boldsymbol{H}} = - c \operatorname{curl} \boldsymbol{E} \quad . \quad . \quad . \quad (5)$$

We shall consider as the simplest particular solution of these differential equations the case of a *plane wave* which propagates itself in the direction of one of the co-ordinates, say in that of the positive x-direction. Then all the field-components are independent of y and z and we get from (5) and (2), since static fields do not come into question for optics:

$$\boldsymbol{E}_x = 0, \ \boldsymbol{H}_x = 0$$

whereas the following differential equations hold for the other components:

$$\epsilon \frac{\partial \boldsymbol{E}_y}{\partial t} = - c \frac{\partial \boldsymbol{H}_z}{\partial x}, \ \epsilon \frac{\partial \boldsymbol{E}_z}{\partial t} = c \frac{\partial \boldsymbol{H}_y}{\partial x},$$

$$\frac{\partial \boldsymbol{H}_y}{\partial t} = c \frac{\partial \boldsymbol{E}_z}{\partial x}, \ \frac{\partial \boldsymbol{H}_z}{\partial t} = - c \frac{\partial \boldsymbol{E}_y}{\partial x}$$

Thus there are two pairs of connected quantities among these four field-components; namely $\boldsymbol{E}_y$ is connected

with $\boldsymbol{H}_z$ and $\boldsymbol{E}_z$ with $\boldsymbol{H}_y$, and the same differential equation holds for each individual component, namely:

$$\frac{\partial^2 \boldsymbol{E}_y}{\partial t^2} = \frac{c^2}{\epsilon} \cdot \frac{\partial^2 \boldsymbol{E}_y}{\partial x^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

So if we set:

$$\frac{c^2}{\epsilon} = q^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

it follows from the general integral already derived in II, § 35, for the differential equation (6) that the most general expression for a plane wave which propagates itself in a homogeneous isotropic medium in the direction of the positive x-axis is:

$$\left.\begin{array}{ll} \boldsymbol{E}_x = 0 & \boldsymbol{H}_x = 0 \\ \boldsymbol{E}_y = \dfrac{1}{\sqrt{\epsilon}} f\left(t - \dfrac{x}{q}\right) & \boldsymbol{H}_y = -\, g\left(t - \dfrac{x}{q}\right) \\ \boldsymbol{E}_z = \dfrac{1}{\sqrt{\epsilon}} g\left(t - \dfrac{x}{q}\right) & \boldsymbol{H}_z = f\left(t - \dfrac{x}{q}\right) \end{array}\right\} \quad . \quad . \quad (8)$$

where f and g represent arbitrary functions of a single argument.

As we see, both field-strengths are perpendicular to the direction of propagation; hence the wave is called "transversal." It resolves into two components which are in general independent of one another, and which lie in the direction of the co-ordinate axes. In the case of each component the electric and the magnetic field-strengths are proportional to one another. Their signs are determined by the theorem that the directions of the electric field-strength, of the magnetic field-strength and of propagation form a right-handed system.

§ 4. If we now propose to ourselves the question as to what is to be measured in this electromagnetic wave and which of its properties can hence be ascertained objectively, we find the answer in the vector of energy-radiation (3) which in the present case reduces to its x-component:

$$\boldsymbol{S}_x = \frac{c}{4\pi}(\boldsymbol{E}_y \boldsymbol{H}_z - \boldsymbol{E}_z \boldsymbol{H}_y) = \frac{q}{4\pi}(f^2 + g^2)$$

Thus in an isotropic body the direction of the energy-radiation coincides with that of the wave-normal x, and the amount of energy radiated in the time dt through a surface F which lies in a wave-plane is :

$$\mathfrak{S}_x \cdot F \cdot dt = \frac{q}{4\pi}(f^2 + g^2) \cdot Fdt \quad . \quad . \quad . \quad (9)$$

Since, however, appreciable effects of radiation always require a finite time, we never measure the radiation vector $\mathfrak{S}_x$ itself, but rather only its time-integral or its mean value in time taken over a sufficiently great interval of time T. Hence if we use the following abbreviations for the mean values :

$$\frac{1}{T}\int_0^T f^2 dt = \bar{f^2}, \quad \frac{1}{T}\int_0^T g^2 dt = \bar{g^2} \quad . \quad . \quad . \quad (10)$$

then the amount of energy radiated through the surface F in unit time is :

$$\frac{q}{4\pi}(\bar{f^2} + \bar{g^2}) \cdot F \quad . \quad . \quad . \quad . \quad . \quad (11)$$

which can be recorded by any instrument that takes up the radiant energy completely, and provided that it is sufficiently sensitive (bolometer, radiometer, thermopile).

After the total radiation of the wave has been measured, its further analysis presents a two-fold problem ; firstly, we must separate the two summands $\bar{f^2}$ and $\bar{g^2}$ from each other ; secondly, we must pass from the mean time values to the functions themselves ; that is, we must investigate the exact form of the wave-functions f and g. For this purpose we require special optical contrivances the theory and action of which we must derive in the sequel. At this stage nothing at all can be stated about them. In particular there is no reason for assigning any sort of periodicity to the functions f and g. Actually there are in optics no waves which have a sharply definite period in

the mathematical sense, such as we have, say, in acoustics. We therefore do best by leaving the question of the form of the waves completely aside for the present, taking it into consideration only when it becomes really necessary. There is only one assumption which we may make from the very outset, namely, that the mean time values of f and g vanish, that is :

$$\bar{f} = 0 \text{ and } \bar{g} = 0 . \quad . \quad . \quad . \quad . \quad (12)$$

For if a wave-function has a mean value different from zero, we can imagine the wave in question to be replaced by another wave for which the conditions (12) are fulfilled, with a statical field superposed on it, the field being characterized by the mean value, which is not equal to zero. The presence of this field can be made manifest by its ponderomotive action on a charged test-body (electron) and can so be separated from the true optical wave.

§ 5. A plane-wave of unlimited cross-section cannot, of course, be realized in nature. Nevertheless we can produce waves which approximate appreciably to the character of plane-waves. For let us imagine a point-like source of light which begins to emit light at a definite moment of time—say when $t = 0$. Then, since the surrounding medium has been assumed to be homogeneous and isotropic, the light will propagate itself uniformly in all directions. The bounding surface which has been reached by the light after a definite interval of time is called a *wave-front*. There is thus a wave-front corresponding to every moment of time, and the whole of the surrounding space is hence filled by the system of successive wave-fronts which enclose one another. In the present case these wave-fronts are obviously spherical surfaces which surround the source of light concentrically, and so a small portion of a sufficiently great spherical surface can be regarded to a sufficient degree of approximation as a plane wave-front or a wave-plane. Its normal is the

corresponding radius of the sphere, and the radiation vector S points in the same direction.

§ 6. Let us now investigate the phenomena that occur when the plane-wave (8) falls on the plane bounding surface of a second isotropic body. We shall take the normal of this bounding plane, the so-called incident normal, as the ξ-axis of a new co-ordinate system directed towards the interior of the second body, whereas the origin 0 of the xyz-system is to be coincident with the origin of the $\xi\eta\zeta$-system. Without reducing the generality of the case we can then make the y-axis and also the η-axis lie in the plane defined by x and ξ, the so-called incident plane, and take this as the plane of Fig. 1. Here all points for which $\xi < 0$ denote the first body (on the left), from which the wave (8) comes, and all points for which $\xi > 0$ denote the second body (on the right); the points for which $\xi = 0$ (the η-axis) constitute the boundary plane. The x-axis is the direction of the ray which comes from the first body—that is, from the left-hand side; it makes the angle θ with the incident normal ξ. The y-axis denotes the wave-plane of the incident ray; this wave-plane is perpendicular to the plane of the figure and also makes the angle θ with the boundary plane. It has been omitted in the figure so as not to multiply the directions to be shown unnecessarily. The z-axis coincides with the ζ-axis and points from the plane of the figure towards the observer.

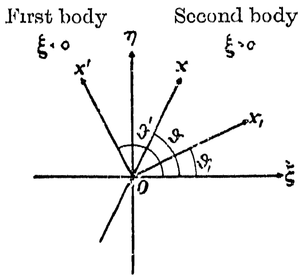

Fig. 1.

We base the solution of the problem before us on the reflection that every system of waves which satisfies the differential equations in the interior of the two bodies and also the boundary conditions, represents a process which is possible in nature.

In order to have the differential equations satisfied in

the second body we imagine a plane-wave in it also, after the model of equations (8), which has the ray-direction x_1 (see Fig. 1), inclined at an angle θ_1 to the ξ-axis, and the wave-plane $y_1 z_1$, where we shall again suppose z_1 to coincide with z and ζ. Then the equations (8) hold for the six field-components $\boldsymbol{E}_{x_1}$, $\boldsymbol{E}_{y_1}$, $\boldsymbol{E}_{z_1}$, $\boldsymbol{H}_{x_1}$, $\boldsymbol{H}_{y_1}$, $\boldsymbol{H}_{z_1}$, except that the co-ordinate x_1 now occurs in place of x on the right-hand side of these equations, the functions f_1 and g_1 replace the wave functions f and g, while the constants ϵ and q are supplanted by the dielectric constant ϵ_1 and, by (7), the velocity of propagation :

$$q_1 = \frac{c}{\sqrt{\epsilon_1}} = q\sqrt{\frac{\epsilon}{\epsilon_1}} \quad . \quad . \quad . \quad . \quad (13)$$

in the second body.

But this assumption does not suffice. For by III, § 6, the boundary conditions require that for $\xi = 0$ the values of the tangential field-components—that is, the quantities $\boldsymbol{E}_\eta$, $\boldsymbol{E}_\zeta$, $\boldsymbol{H}_\eta$, $\boldsymbol{H}_\zeta$—are coincident in both bodies. This gives four equations connecting the wave-functions; to satisfy them, however, we have only the two functions f_1 and g_1 available, since the functions f and g are initially given. To generalize our initial assumption still further, therefore, we assume a second wave in the first body; this wave is, of course, also represented by the equations (8), except that it has a different ray-direction x', which we shall assume to make an angle θ' with the ξ-axis (see Fig. 1), and has the wave-plane $y'z'$, where again $z' = z$. The six field-components $\boldsymbol{E}_{x'}$, $\boldsymbol{E}_{y'}$, $\boldsymbol{E}_{z'}$, $\boldsymbol{H}_{x'}$, $\boldsymbol{H}_{y'}$, $\boldsymbol{H}_{z'}$ are given by the equations (8), if we substitute in them the wave-functions f' and g' and the co-ordinate x', the constants ϵ and q remaining the same.

We have now approximately generalized our assumption for the interior of the two bodies, and can proceed to set up the boundary conditions. In the first body there is an electromagnetic field which results from the superposition of the two plane waves that we have assumed. Hence, remembering that the field-components $\boldsymbol{E}_x$, $\boldsymbol{H}_x$,

$\boldsymbol{E}_{x'}$, $\boldsymbol{H}_{x'}$ vanish, we get for the field-components which interest us in the first body:

$$\boldsymbol{E}_\eta = \boldsymbol{E}_y \cos\theta + \boldsymbol{E}_{y'} \cos\theta' = \frac{\cos\theta}{\sqrt{\epsilon}} \cdot f + \frac{\cos\theta'}{\sqrt{\epsilon}} \cdot f'$$

$$\boldsymbol{E}_\zeta = \boldsymbol{E}_z + \boldsymbol{E}_{z'} = \frac{1}{\sqrt{\epsilon}} \cdot g + \frac{1}{\sqrt{\epsilon}} \cdot g'$$

$$\boldsymbol{H}_\eta = \boldsymbol{H}_y \cos\theta + \boldsymbol{H}_{y'} \cos\theta' = -\cos\theta \,.\, g - \cos\theta' \,.\, g'$$

$$\boldsymbol{H}_\zeta = \boldsymbol{H}_z + \boldsymbol{H}_{z'} = f + f'$$

On the other hand, we obtain for the second body ($\xi > 0$), remembering that the field-components $\boldsymbol{E}_{x_1}$ and $\boldsymbol{H}_{x_1}$ vanish:

$$\boldsymbol{E}_\eta = \boldsymbol{E}_{y1} \cos\theta_1 = \frac{\cos\theta_1}{\sqrt{\epsilon_1}} f_1$$

$$\boldsymbol{E}_\zeta = \boldsymbol{E}_{z1} = \frac{1}{\sqrt{\epsilon_1}} g_1$$

$$\boldsymbol{H}_\eta = \boldsymbol{H}_{y1} \,.\, \cos\theta_1 = -\cos\theta_1 \,.\, g_1$$

$$\boldsymbol{H}_\zeta = \boldsymbol{H}_{z1} = f_1.$$

Hence if we use the abbreviation:

$$\sqrt{\frac{\epsilon_1}{\epsilon}} = \frac{g}{g_1} = n \quad . \quad . \quad . \quad . \quad . \quad (14)$$

we must have for the boundary plane $\xi = 0$:

$$\cos\theta \,.f + \cos\theta' .\, f' = \frac{\cos\theta_1}{n} \cdot f_1$$

$$g + g' = \frac{g_1}{n}$$

$$\cos\theta \,.\, g + \cos\theta' \,.\, g' = \cos\theta_1 \,.\, g_1$$

$$f + f' = f_1$$

These four equations comprehend all the details of the theory of reflection and refraction. As we see, they fall into two groups, one of which contains only the f-waves, and the other only the g-waves. Thus these two kinds of

waves behave quite independently of one another; each obeys its own laws.

§ 7. By means of the last four equations we first calculate the unknown wave-functions f', f_1, g', g_1 from the given wave-functions f and g. We get:

$$f' = \frac{n\cos\theta - \cos\theta_1}{\cos\theta_1 - n\cos\theta'} \cdot f = \mu \, . f \quad . \quad . \quad (15)$$

$$f_1 = \frac{n(\cos\theta - \cos\theta')}{\cos\theta_1 - n\cos\theta'} \cdot f = \mu_1 \cdot f \quad . \quad . \quad (16)$$

$$g' = \frac{\cos\theta - n\cos\theta_1}{n\cos\theta_1 - \cos\theta'} \cdot g = \sigma \, . g \quad . \quad . \quad (17)$$

$$g_1 = \frac{n(\cos\theta - \cos\theta')}{n\cos\theta_1 - \cos\theta'} \cdot g = \sigma_1 \, . g \quad . \quad . \quad (18)$$

As for the arguments of these functions, we have:

$$t - \frac{x}{q} \text{ in } f \text{ and } g$$

$$t - \frac{x_1}{q_1} \text{ in } f_1 \text{ and } g_1$$

$$t - \frac{x'}{q} \text{ in } f' \text{ and } g'$$

And $\xi = 0$ everywhere, so that in transforming to the co-ordinates ξ, η, ζ we have:

$$x = \eta \sin\theta, \;\; x_1 = \eta \sin\theta_1, \;\; x' = \eta \sin\theta'$$

which makes the arguments assume the values:

$$t - \frac{\eta\sin\theta}{q}, \;\; t - \frac{\eta\sin\theta_1}{q_1}, \;\; t - \frac{\eta\sin\theta'}{q}.$$

Since the functional equations (15) to (18) must be satisfied for all times t, and for all points η of the boundary surface, it follows that these three arguments must be equal to one another—as can also be seen directly if we differentiate one of the functional equations by parts first

with respect to t and then with respect to η, and divide the resulting equations by one another. We get:

$$\frac{\sin\theta}{q} = \frac{\sin\theta_1}{q_1} = \frac{\sin\theta'}{q} \quad . \quad . \quad . \quad . \quad . \qquad (19)$$

and hence arrive at the *law of refraction*:

$$\frac{\sin\theta}{\sin\theta_1} = \frac{q}{q_1} = n = \sqrt{\frac{\epsilon_1}{\epsilon}} \quad . \quad . \quad . \quad . \qquad (20)$$

and the *law of reflection*:

$$\theta' = \pi - \theta \quad . \quad . \quad . \quad . \quad . \quad . \qquad (21)$$

If we call the angle which the reflected ray makes with the reversed incident normal the angle of reflection, then the angle of reflection is equal to the angle of incidence.

§ 8. Snell's law of refraction (20), which states that the ratio of the sine of the angle of incidence θ to the sine of the angle of refraction θ_1 is equal to the refractive index n of the second body with respect to the first or to the ratio of the velocities of propagation q and q_1, has been accurately confirmed by innumerable measurements. The refractive index of a substance is usually referred to air as the first substance. Thus the refractive index of water is equal to 1·3, that of glass to 1·5. We then obtain the refractive index of a substance with respect to any other substance by writing down the ratio of their refractive indices with respect to air. If we exchange the substances, the refractive index assumes the reciprocal of its previous value. Accordingly, the refractive index with respect to a vacuum—the so-called "absolute" refractive index—is the product of the refractive index with respect to air and of the absolute refractive index of air namely 1·0003; as we see, its value differs in most cases only inappreciably from the ordinary refractive index.

If we allow the angle of incidence θ to vary from 0 (normal incidence) to $\frac{\pi}{2}$ (grazing incidence), the angle

of refraction θ_1 increases from 0 to $\sin^{-1}\frac{1}{n}$ (limiting angle). But there is a point of fundamental importance which must not be overlooked. It is only when $n>1$, or if, as we say, the second substance is optically denser than the first, that the limiting angle is real. Then the angle of refraction θ_1 is always smaller than the angle of incidence θ—that is, the ray is bent towards the incident normal by the refraction, and the limiting angle denotes the greatest value which the refractive index can assume at all. But if $n<1$—that is, if we exchange the two substances with each other—the angles of incidence and refraction also exchange their rôles, and the angle of refraction becomes greater than the angle of incidence; it attains the value $\frac{\pi}{2}$ only when the angle of incidence has reached the value of the limiting angle. If the angle of refraction is allowed to go beyond the limiting angle, then (20) leads to an imaginary value for the angle of refraction, and the solution which we have found for the problem of refraction becomes meaningless. As there is nothing to prevent our giving the angle of incidence any arbitrary value between 0 and $\frac{\pi}{2}$, a special question arises here, which we shall, however, deal with on a later occasion (§ 12); for the present we shall restrict ourselves to considering those cases for which the law of refraction yields a real value for the angle of refraction θ_1.

§ 9. But the electromagnetic theory of the refraction of light states more than that the refractive index is independent of the value of the angle of incidence; it also tells us the value of the refractive index. For by (20) this is equal to the square root of the ratio of the dielectric constants, or, if we take as our basis the absolute refractive index:

$$n = \sqrt{\epsilon_1}. \qquad (22)$$

If we compare this relationship with observed facts, we

find, in general, crass disagreement. For example, for water $n = 1{\cdot}3$, while $\epsilon_1 = 80$. But even apart from this the fact that (22) cannot be generally valid follows from the fact that by definition the dielectric constant ϵ is independent of the form of the wave-functions f and g, whereas the refractive index n, in the case of *all* substances, depends more or less markedly on the form of the light-waves, that is, on the colour of the light. This phenomenon, dispersion, long constituted a serious obstacle to the acceptance of Maxwell's theory. If we wish to take adequate account of it in the theory here described, nothing remains but to conclude that the fundamental assumption which was introduced at the beginning of this chapter into the field-equations for the optics of homogeneous and isotropic bodies—namely, the relation (4), which states that the electric induction is proportional to the electric intensity of field—does not in general correspond with reality in the case of rapid optical vibrations. To obtain a satisfactory theory of dispersion we shall therefore have to replace this relationship by one that is more general. This will be done in the third part of the present volume, where it will be found that this generalization will have to be based on the circumstance that in the case of refined optical phenomena in nature the assumption that matter is absolutely continuous and homogeneous is no longer justified, but must be modified by the introduction of characteristic structural properties to a certain extent.

If this view is correct, an important significance will still have to be attached to the relation (22)—namely, that of a limiting law which is the better fulfilled the less the dispersion makes itself observed. If we carry out a test in this direction, the relationship in question is found to be definitely confirmed. For the substances which disperse least are gases, and the earliest measurements, by L. Boltzmann, have accurately confirmed the formula (22) in their case. A particularly noteworthy feature is the exact quantitative parallelism between the

dependence of the refractive index and of the dielectric constant on the pressure in the case of gases, and this occurs in the sense of equation (22). Hence, with reference to this equation, we are right in speaking of a far-reaching confirmation of the electromagnetic theory within the admissible range of application.

§ 9*a*. But besides giving the *directions* of the reflected and the refracted rays, the theory also gives the *form* of the reflected and the refracted rays, by demanding that the wave-functions in question shall be proportional to the corresponding wave-functions of the incident wave. If in the formulæ (15) to (18) we replace the refractive indices n according to (20) by the angles θ and θ_1, and the angle θ' by $\pi-\theta$, the constants of proportionality assume the following values :

for the reflected wave (f', g') :

$$\mu = \frac{\tan(\theta-\theta_1)}{\tan(\theta+\theta_1)}, \quad \sigma = \frac{\sin(\theta-\theta_1)}{\sin(\theta+\theta_1)} \quad . \quad . \quad (23)$$

for the refracted wave (f_1, g_1) :

$$\mu_1 = \frac{\sin 2\theta}{\sin(\theta+\theta_1)\cos(\theta-\theta_1)}, \quad \sigma_1 = \frac{\sin 2\theta}{\sin(\theta+\theta_1)} \quad (24)$$

According to these formulæ (known as Fresnel's formulæ) there is a fundamental difference between the two wave-functions f and g, which corresponds to the physical circumstance that, according to (8), in the case of the f-wave the electric intensity of field lies in the plane of incidence, whereas in the case of the g-wave the electric intensity of field is in a direction perpendicular to the plane of incidence. The coefficients μ correspond to the former, the coefficients σ to the latter.

To test the theory we have to measure the radiant energy. Let us first consider the *reflected wave*. From equation (11), using (15) and (17), we get for the ratio of the intensity of radiation of the reflected wave to

the intensity of radiation of the incident wave, that is, the "reflection coefficient":

$$\frac{\overline{f'^2} + \overline{g'^2}}{\overline{f^2} + \overline{g^2}} = \rho = \frac{\mu^2\overline{f^2} + \sigma^2\overline{g^2}}{\overline{f^2} + \overline{g^2}}$$

As we see, the reflection coefficient ρ lies between the values μ^2 and σ^2 which correspond with the limiting cases where one of the two waves f and g is vanishingly small. In general, by measuring ρ we find for the ratio of the intensities of radiation of the waves f and g:

$$\overline{f^2} : \overline{g^2} = \frac{\sigma^2 - \rho}{\rho - \mu^2} \quad . \quad . \quad . \quad . \quad . \quad (25)$$

This relation may be tested experimentally by measuring the reflection coefficient ρ of a definite incident wave (f, g) for different angles of incidence θ and, after introducing the values of μ and σ calculated from (23), investigating whether the same value comes out for the intensity ratio (25) for all angles of incidence. This has been confirmed experimentally in all cases where the surface of the body used has been sufficiently smooth and where sufficient precautions have been taken to avoid the chemical impurities that often contaminate surfaces, including those due to the use of polishing agents.

Light derived from a body which has been made to glow by having its temperature increased is called "natural light." For this, measurement gives, as we should expect, $\overline{f^2} = \overline{g^2}$, and hence:

$$\rho = \frac{\mu^2 + \sigma^2}{2} \quad . \quad . \quad . \quad . \quad . \quad (26)$$

The dependence of the reflection coefficient ρ on the angle of incidence θ is obtained from (23). For normal incidence we have, since θ and θ_1 are infinitely small:

$$\rho = \left(\frac{\theta - \theta_1}{\theta + \theta_1}\right)^2 = \left(\frac{n - 1}{n + 1}\right)^2 \quad . \quad . \quad . \quad (27)$$

The reflection is the greater, the more the refractive index n differs from the value unity, that is, the more the two bodies differ optically. For grazing incidence, $\theta = \frac{\pi}{2}$, or, for the limiting angle, $\theta_1 = \frac{\pi}{2}$, the reflection coefficient attains its maximum value 1; in that case no radiation penetrates into the second body at all.

But the inferences to be drawn from the theory and its possible applications go much further. For if we consider the constitution of the reflected radiation we find that if the incident light consisted of natural light, the reflected light does not, but that rather, by (15) and (17):

$$\overline{f'^2} : \overline{g'^2} = \frac{\mu^2}{\sigma^2} = \frac{\cos^2(\theta + \theta_1)}{\cos^2(\theta - \theta_1)} \quad . \quad . \quad . \quad . \quad (28)$$

It is only for normal and grazing incidence (or in the case of the limiting angle, respectively) that this ratio is equal to 1 and that the reflected radiation consists of natural light—which is obvious at once, since in the first case, the incident plane is indeterminate, and hence there is no physical difference between the waves f' and g', whereas in the latter case *all* the light is reflected. But in general the two components have different intensities—that is, the reflected radiation is "polarized," the intensity of the f-wave, whose electric intensity of field lies in the incident plane, being smaller than that of the g-wave. In fact, for the case:

$$\theta + \theta_1 = \frac{\pi}{2}, \text{ or } \tan\theta = n \quad . \quad . \quad . \quad . \quad (29)$$

that is, when the reflected ray is perpendicular to the refracted ray, the quotient (28) is equal to zero, so that the f-wave is absent altogether in the reflected light, and the g-wave alone remains. The radiation is then said to be "completely" polarized "linearly," because in the present case the electric and the magnetic intensity of field of the wave have perfectly definite directions. The angle of incidence defined by (29) is called the "angle of polarization," the plane of incidence the "plane of polariza-

tion" and the reflecting body (mirror) is called the "polarizer." If we call the plane defined by the direction of the ray and the electric intensity of field the "plane of vibration" of the ray, then the plane of vibration in the light reflected at the angle of polarization (29) is constant and perpendicular to the plane of incidence (Brewster's Law). This law can be directly tested experimentally by allowing the linearly polarized ray to be reflected from a second body of the same substance—namely, at the polarizing angle—but in such a way that at the second reflection the incident ray assumes the rôle of the f-wave—that is, so that the plane of vibration coincides with the new plane of incidence. This occurs when the plane of incidence at the second reflection is perpendicular to the plane of incidence at the first reflection. In this case no light is reflected at all (Malus's mirror experiment). The second reflecting body then acts as an "analyser."

Thus the reflection at the angle of polarization forms, quite generally, a means of measuring the two wave-components f and g present in any given radiation. For by measuring the intensity of the whole reflected light we obtain, since f' vanishes:

$$\overline{g'^2} = \sigma^2 \overline{g^2} = \left(\frac{n^2 - 1}{n^2 + 1}\right)^2 \overline{g^2} \quad . \quad . \quad . \quad . \quad (30)$$

and from this the intensity of the g-wave; and in the same way, by reflection at an incident plane turned through $\frac{\pi}{2}$, the intensity of the f-wave is obtained.

When the reflection takes place at any arbitrary angle of incidence θ the reflected ray is only partially polarized. The intensities of the two wave-components $\overline{f'^2}$ and $\overline{g'^2}$ are then obtained by multiplying the corresponding components $\overline{f^2}$ and $\overline{g^2}$ in the incident wave by the corresponding reflection coefficients. For the light whose vibrations are in the plane of incidence (f-wave) we have:

$$\rho_{\|} = \mu^2 = \frac{\tan^2 (\theta - \theta_1)}{\tan^2 (\theta + \theta_1)} \quad . \quad . \quad . \quad . \quad (31)$$

For the light whose vibrations are perpendicular to the incident plane (g-wave) we have :

$$\rho_{\perp} = \sigma^2 = \frac{\sin^2(\theta - \theta_1)}{\sin^2(\theta + \theta_1)} \quad . \quad . \quad . \quad . \quad (32)$$

§ 10. Exactly similar considerations apply to the radiation *transmitted* by the boundary surface. They can be reduced to those of the preceding paragraph by recollecting that according to the principle of the conservation of energy the transmitted radiation is equal to the excess of incident radiation over reflected radiation. For the boundary conditions (cf. III, § 6) provide that no energy of radiation still becomes lost. If we take as the definition of the "coefficient of transmissibility" the fraction of the incident radiation which passes through the boundary surface into the second body, the transmission coefficient for the light which vibrates in the plane of incidence (f-wave) is :

$$1 - \rho_{\parallel} = 1 - \mu^2 = \frac{\sin 2\theta \,.\, \sin 2\theta_1}{\sin^2(\theta + \theta_1)\cos^2(\theta - \theta_1)} \quad . \quad (33)$$

whereas for the light which vibrates perpendicularly to the incident plane (g-wave) :

$$1 - \rho_{\perp} = 1 - \sigma^2 = \frac{\sin 2\theta \,.\, \sin 2\theta_1}{\sin^2(\theta + \theta_1)} \quad . \quad . \quad (34)$$

Of course, we can also calculate the transmission coefficient directly from the ratio of the intensity of the wave f_1 or g_1 which penetrates into the second body to that of the incident wave f or g. Now we may not, by referring say to (16) and (18), set the transmission coefficient in question equal to ${\mu_1}^2$ or ${\sigma_1}^2$, but must revert to the expression (11) which gives the energy radiated per unit of time through a surface F which lies in a wave-plane. This is clear if we consider that in the passage into the second body both the velocity of propagation q and also the extent of the surface F, which represents the normal

cross-section of a selected cylindrical beam of rays, change. A simple geometrical reflection shows that :

$$\frac{F}{\cos\theta} = \frac{F_1}{\cos\theta_1} \quad . \quad . \quad . \quad . \quad . \quad . \quad (35)$$

For this is the size of the surface which is cut out of the limiting plane ($\xi = 0$) on the one hand by the cylinder of the incident rays, on the other hand by the cylinder of transmitted rays which results from it.

Hence we get as the ratio of the transmitted energy of radiation to the incident energy for the f-wave by (11), (20), (35) and (16) :

$$\frac{q_1\overline{f_1^2}F_1}{q\overline{f^2}F} = \frac{\sin\theta_1}{\sin\theta} \cdot \mu_1^2 \cdot \frac{\cos\theta_1}{\cos\theta}$$

and correspondingly for the g-wave :

$$\frac{q_1\overline{g_1^2}F_1}{q\overline{g^2}F} = \frac{\sin\theta_1}{\sin\theta} \cdot \sigma_1^2 \cdot \frac{\cos\theta_1}{\cos\theta}$$

and these expressions, by (24), agree with (33) and (34). The ratio of the two components of radiation in the transmitted radiation is :

$$\frac{\overline{f_1^2}}{\overline{g_1^2}} = \frac{\mu_1^2}{\sigma_1^2} = \frac{1}{\cos^2(\theta - \theta_1)} \quad . \quad . \quad . \quad . \quad (36)$$

Since this expression can become neither zero nor infinite, the transmitted radiation is never completely, but only partially polarized in the case of incident natural light; and the component whose vibrations lie in the plane of incidence is the more intense; this is the reverse of what happens in the case of reflected radiation.

Finally we must refer to a relationship which expresses a general law of optics. Both the equation (20) of the law of refraction and the values of the reflection and the transmission coefficients remain unaltered if we exchange the two bodies and simultaneously the angles θ and θ_1—that is, if we allow the wave to fall in the reverse direction from the second body on to the boundary surface of the

first body. In other words, light does not only pursue the reversed path, but is also divided in the same ratio into reflected and transmitted radiation. A boundary surface of two bodies reflects equally well in both directions and also transmits equally well. This law is an application of a law of reciprocity first enunciated by Helmholtz, which states that the weakening of intensity which a definite ray of light undergoes in its passage through arbitrary different bodies in consequence of reflection, refraction, diffusion and absorption is independent of the direction in which the ray travels.

§ 11. Now that we have seen that the components f and g of a given plane wave (8) which vibrate perpendicularly to one another can be separated from one another by measurement and have seen how this can be done, the further question arises whether the values of the two intensities $\overline{f^2}$ and $\overline{g^2}$ also determine the intensity of the component of radiation which vibrates in another direction perpendicular to the direction of propagation, analogously to the way in which two mutually perpendicular components of a force also determine the component in any other direction. We shall find that this is not the case and shall see wherein the characteristic difference lies.

To decide this question, we again consider the most general case (8) of a plane wave, but for convenience of notation, we shall take the z-axis as the direction of propagation instead of the x-axis, so that the xy-plane becomes the wave-plane of the radiation. Let the f-wave vibrate in the x-direction, and the g-wave in the y-direction. If we now replace the axes x and y by two other mutually perpendicular axes x' and y' that lie in the same plane and that make an angle ϕ with the former axes, then since the wave-functions f' and g' by (8) transform like the components of the field-intensity, they become, when referred to these new axes :

$$f' = f\cos\phi + g\sin\phi \quad . \quad . \quad . \quad . \quad (37)$$

$$g' = -f\sin\phi + g\cos\phi \quad . \quad . \quad . \quad (38)$$

and hence the corresponding intensities :

$$\overline{f'^2} = \overline{f^2}\cos^2\phi + 2\overline{fg}\sin\phi\cos\phi + \overline{g^2}\sin^2\phi$$
$$\overline{g'^2} = \overline{f'^2}\sin^2\phi - 2\overline{fg}\sin\phi\cos\phi + \overline{g^2}\cos^2\phi.$$

Adding these expressions we get :

$$\overline{f'^2} + \overline{g'^2} = \overline{f^2} + \overline{g^2}$$

that is, we always get the total intensity of radiation of the whole wave if we add together the intensities of two components which vibrate in perpendicular directions; this is, indeed, obvious from the fact that the original xy-system can be chosen quite arbitrarily. But the intensity of any single component, say that whose direction of vibration makes the angle ϕ with the x-axis, does not alone depend on the intensities $\overline{f^2}$ and $\overline{g^2}$, but also on a third quantity, $\overline{fg}$. So if we use the abbreviations :

$$\overline{f^2} = A,\ \overline{g^2} = B,\ \overline{fg} = C \quad . \quad . \quad . \quad (39)$$

then the component of the intensity of radiation vibrating at an azimuth ϕ assumes the value :

$$\overline{f'^2} = J_\phi = A\cos^2\phi + B\sin^2\phi + 2C\sin\phi\cos\phi \quad . \quad (40)$$

Accordingly, the intensity of radiation is not represented by a vector, but by a tensor of the second degree (II, § 13, § 20)—namely, by a plane tensor which has only three components and which is characterized by not having the quantity (40) dependent on the choice of the co-ordinate system. The component C can have any value between $-\sqrt{AB}$ and $+\sqrt{AB}$, but it cannot overstep this range, since J is positive for all values of J.

For one of the limiting cases, $C = \pm\sqrt{AB}$, J by (40) becomes a perfect square :

$$J = (\sqrt{A}\,.\cos\phi \pm \sqrt{B}\,.\sin\phi)^2 \quad . \quad . \quad (41)$$

So $J = \overline{f'^2}$ vanishes for the azimuth $\tan\phi = \mp\sqrt{\dfrac{A}{B}}$ and the whole radiation reduces itself to a single component

$\overline{g'^2}$ which vibrates in the perpendicular azimuth. In this case the light is linearly polarized. From $\overline{f'^2} = 0$ it follows that $f' = 0$ and by (37):

$$\frac{f}{g} = -\tan\phi = \pm\sqrt{\frac{A}{B}} \quad . \quad . \quad . \quad . \quad (42)$$

that is, the two wave-functions f and g bear a fixed ratio to each other which is independent of their argument. In such a case, where the one wave-function is completely determined by the other, the two corresponding components of radiation are said to be "completely coherent." Thus linear polarization always implies complete coherence of the two components into which the polarized ray can be resolved. It follows directly from the relationships (15) to (18) that linearly polarized light remains linearly polarized after refraction and reflection, since then the ratios $f_1 : g_1$ and $f' : g'$ are also constant.

In general the light is only partially polarized, so that the two components are partly non-coherent. Natural light forms the opposite limiting case, for which the intensity of radiation J_ϕ is quite independent of the azimuth of vibration ϕ. By (40) this not only leads to $A = B$, but also to $C = 0$. As may easily be seen, the latter condition is always fulfilled in actual fact, if the two vibration components f and g are completely non-coherent—that is, fully independent of each other. For this means that to a given value of f there belongs not one definite value of g, but an enormous number of values of g, and that the mean of all these values, $\bar{g}$, is independent of the value assumed by f. From this there follows as the mathematical expression for the complete non-coherence of f and g:

$$\overline{fg} = \bar{f} \cdot \bar{g} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (43)$$

and by (39) and (12):

$$C = \overline{fg} = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (44)$$

Since the functions of f and g depend uniquely on their

common argument, a necessary preliminary condition for the non-coherence of f and g is, that corresponding to a definite value of f or g there must be an enormous number of values of the argument, or that the value in question of the wave-function must recur enormously often in the course of time.

Although the relation (44) has been shown to be a necessary consequence of the complete non-coherence of the two wave-functions f and g, this relation is by no means sufficient to ensure non-coherence. Indeed, we shall soon see (§ 20) that the components of f and g can actually be completely coherent even when the equation (44) is simultaneously fulfilled. Of course f and g are not proportional to one another, as in (42)—that is, the light is not linearly polarized—but nevertheless it is completely polarized, since the one component is completely determined at every moment by the other component.

If we now again turn to consider the general case, we easily see that the intensity of radiation J_ϕ has by (40) a maximum and a minimum, both of which result from the equation :

$$\frac{dJ}{d\phi} = 0$$

and occur for the angle ϕ, which is obtained from :

$$\tan 2\phi = \frac{2C}{A - B} \quad . \quad . \quad . \quad . \quad (45)$$

These two directions, which are mutually perpendicular, are called the principal planes of vibration of the ray; the corresponding intensities—that is, the maximum and minimum of J :

$$J = \frac{A + B}{2} \pm \frac{1}{2}\sqrt{(A - B)^2 + 4C^2} \quad . \quad . \quad (46)$$

are called the principal intensities of radiation.

For linearly polarized light we have, as we saw, $C = \pm\sqrt{AB}$. Then by (46) the maximum intensity is :

$$J = A + B$$

and the minimum intensity is :

$$J = 0$$

For natural light $A = B$ and $C = 0$; hence, by (46), $J = A = B$, whereas the principal planes of vibration become indeterminate, by (45).

If the co-ordinate axes x and y are made to lie in the directions of the principal planes of vibration, the equation (45) is satisfied by $\phi = 0$ and $\phi = \frac{\pi}{2}$. Consequently $C = 0$ in that case, and the component of the intensity of radiation in any arbitrary direction ϕ becomes, by (40) :

$$J_\phi = A \cos^2 \phi + B \sin^2 \phi \quad . \quad . \quad . \quad (47)$$

So A and B are the principal intensities, in agreement with (46).

Since in general the intensity of radiation J_ϕ depends, according to (40), on three constants, it follows that to characterize the state of polarization of the radiation it is not sufficient to measure the two components $J_0 = A$ and $J_{\frac{\pi}{2}} = B$. Actually, for example, when $A = B$, the light might be natural light or light linearly polarized at an azimuth of 45°, or it might be composed of some mixture of natural and linearly polarized light. To determine the third tensor-component $\overline{fg} = C$, it is thus necessary to make a third measurement, say of the component which vibrates at the angle $\frac{\pi}{4}$:

$$J_{\frac{\pi}{4}} = \frac{A + B}{2} + C \quad . \quad . \quad . \quad . \quad (48)$$

Then A, B and C are fully determined.

§ 12. *Total Reflection.* Having disposed of the questions relating to the phenomena of reflection and refraction, we have still to deal with the case which we deferred provisionally at the conclusion of § 8 because the solution we found was inapplicable to it. This is the case where

$\sin\theta > n$ $(n < 1)$ because the angle of refraction θ_1 then becomes imaginary, by (20).

It is obvious that it is not sufficient to say here that since the angle of refraction is imaginary, no refraction can occur, and consequently all the light is reflected. For the equations of the problem are by no means satisfied by setting the wave-functions of the refracted wave equal to zero and those of the reflected wave equal to those of the incident wave. It is impossible to evade the task of finding expressions for the wave-functions which fulfil all the conditions demanded by the theory. On the other hand, we must bear in mind that a complex solution of an equation is also a solution, and that even if it has no real significance, it may easily serve as a guide in finding a real solution.

In fact, every equation between complex quantities resolves into two equations between only real quantities. If we remember also that all the equations between the field-strengths of the incident, reflected and refracted waves—both those which hold in the interior of the body and also those which express the boundary conditions, are linear and homogeneous with respect to the field-strengths and have real coefficients—it follows that if all the equations are satisfied by certain complex values of the field-strengths, the real parts of these complex values also satisfy these equations and hence represent a real physical solution of the problem. Hence we immediately derive a real solution from every arbitrary complex solution merely by omitting the purely imaginary parts from the complex values of the field-intensities and retaining only the real parts. In putting this idea into practice we reflect that since the angle of refraction θ_1 is complex, the coefficients μ and σ and the wave-functions of the refracted and reflected waves will also assume complex values. For the sake of complete generality we shall also assume the same for the incident wave; that is, we now take f and g to stand for two arbitrary complex functions of a single complex argument. If we denote

the real part of any complex quantity by prefixing an $\boldsymbol{R}$, then we get for the real field-components of the *incident* wave, from (8), the expressions:

$$\left.\begin{array}{ll} \boldsymbol{E}_x = 0 & \boldsymbol{H}_x = 0 \\ \boldsymbol{E}_y = \dfrac{1}{\sqrt{\epsilon}} \boldsymbol{R} f & \boldsymbol{H}_y = -\boldsymbol{R} g \\ \boldsymbol{E}_z = \dfrac{1}{\sqrt{\epsilon}} \boldsymbol{R} g & \boldsymbol{H}_z = \boldsymbol{R} f \end{array}\right\} \quad . \quad . \quad (49)$$

where the quantity $t - \dfrac{x}{q}$ is again to be put as the argument of f and g. As hitherto, x is the wave-normal, whereas the y-axis again lies in the incident plane and the z-axis is perpendicular to it.

As for the *refracted* wave its wave-normal x_1 is complex. Hence we shall refer its real field-components right from the outset to the real co-ordinate system ξ, η, ζ, which is determined by the incident normal ξ and the plane of incidence $\xi\eta$ (Fig. 1). For this purpose we shall first form the complex field-strengths according to the equations:

$$\boldsymbol{E}_\xi = \boldsymbol{E}_{x_1} . \cos\theta_1 - \boldsymbol{E}_{y_1} . \sin\theta_1$$
$$\boldsymbol{E}_\eta = \boldsymbol{E}_{x_1} . \sin\theta_1 + \boldsymbol{E}_{y_1} . \cos\theta_1$$

and so forth, and replace the components $\boldsymbol{E}_{x_1}$, $\boldsymbol{E}_{y_1}$, $\boldsymbol{E}_{z_1}$ . . . in it by the wave-functions f_1, g_1 after the model of (8), and then replace these in turn, by (16) and (18), by the wave-functions f and g of the incident wave. We then get the following expressions for the real field-components of the refracted wave:

$$\left.\begin{array}{ll} \boldsymbol{E}_\xi = -\dfrac{1}{\sqrt{\epsilon_1}} \cdot \boldsymbol{R}\mu_1 f \sin\theta_1 & \boldsymbol{H}_\xi = \boldsymbol{R}\sigma_1 g \sin\theta_1 \\ \boldsymbol{E}_\eta = \dfrac{1}{\sqrt{\epsilon_1}} \cdot \boldsymbol{R}\mu_1 f \cos\theta_1 & \boldsymbol{H}_\eta = -\boldsymbol{R}\sigma_1 g \cos\theta_1 \\ \boldsymbol{E}_\zeta = \dfrac{1}{\sqrt{\epsilon_1}} \cdot \boldsymbol{R}\sigma_1 g & \boldsymbol{H}_\zeta = \boldsymbol{R}\mu_1 f \end{array}\right\} \quad . \quad . \quad (50)$$

Here we have to insert $t - \frac{x_1}{q_1}$ everywhere as the argument of f and g, in that:

$$x_1 = \xi \cos \theta_1 + \eta \sin \theta_1 \quad . \quad . \quad . \quad . \quad (51)$$

Finally, for the *reflected* wave, whose normal x' is real, we obtain, by (8), (15) and (17), the real field-components:

$$\left.\begin{array}{ll} E_{x'} = 0 & H_{x'} = 0 \\ E_{y'} = \frac{1}{\sqrt{\epsilon}} R\mu f & H_{y'} = -R\sigma g \\ E_{z'} = \frac{1}{\sqrt{\epsilon}} R\sigma g & H_{z'} = R\mu f \end{array}\right\} \quad . \quad . \quad (52)$$

with the argument $t - \frac{x'}{q}$ of f and g.

§ 13. Although the preceding equations solve the problem of total reflection in principle, the theory cannot be directly applied to the general case of arbitrary wave-functions because it is not possible to specify the real part of a complex expression so long as its form is not known. For this reason we are compelled, if we wish to arrive at real expressions, first to consider particular solutions of the problem, and then to dispose of the general case by appropriately combining these particular solutions.

In choosing the particular solutions that can be used in optics, we bear in mind that, to represent natural light, we must consider, as we saw in § 11, only such wave-functions as can assume each one of its values enormously often. Hence we choose as the simplest particular solution a simple periodic function by setting:

$$f(t) = e^{i\omega t} = \cos \omega t + i \sin \omega t \quad . \quad . \quad . \quad (53)$$

and we treat the two waves f and g separately by first taking:

$$g(t) = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (54)$$

Then, by (49), we get for the *incident* wave :

$$\left.\begin{array}{ll} \boldsymbol{E}_x = 0 & \boldsymbol{H}_x = 0 \\ \boldsymbol{E}_y = \dfrac{1}{\sqrt{\epsilon}} \cos \omega\left(t - \dfrac{x}{q}\right) & \boldsymbol{H}_y = 0 \\ \boldsymbol{E}_z = 0 & \boldsymbol{H}_z = \cos \omega\left(t - \dfrac{x}{q}\right) \end{array}\right\} \quad . \quad . \quad (55)$$

This is a special sine wave of "radian frequency" ω and "revolution" frequency or vibration number $\nu = \dfrac{\omega}{2\pi}$ and wave-length $\lambda = \dfrac{2\pi q}{\omega}$ (II, § 40).

To find the real field-components for the *refracted* wave from (50) we first calculate the complex quantities $\cos \theta_1$ and μ_1. For the former we get from (20) :

$$\cos \theta_1 = \pm i\theta'$$

where we set the positive real quantity :

$$\sqrt{\frac{\sin^2 \theta}{n^2} - 1} = \theta' . \quad . \quad . \quad . \quad . \quad (56)$$

(which is not to be confused with the angle of reflection in (21)). The coefficient μ_1 is given by (24). We resolve it into its real and its imaginary parts by writing :

$$\mu_1 = \frac{2 \sin \theta \cos \theta}{\sin \theta \cos \theta + \sin \theta_1 \cos \theta_1}$$

$$= \frac{2 \sin^2 \theta \cos^2 \theta \mp i \, . \, 2 \sin \theta \cos \theta \, . \, \theta' \, . \sin \theta_1}{\sin^2 \theta \cos^2 \theta + \theta'^2 \sin^2 \theta_1}$$

or, if we use the abbreviation :

$$\frac{\theta' \sin \theta_1}{\sin \theta \cos \theta} = \tan \frac{\delta}{2}, \ (0 < \delta < \pi) \ . \quad . \quad . \quad (57)$$

$$\mu_1 = 2 \cos^2 \frac{\delta}{2} \mp i \, . \, 2 \sin \frac{\delta}{2} \cos \frac{\delta}{2} = 2 \cos \frac{\delta}{2} e^{\mp i \frac{\delta}{2}}$$

The choice of sign is determined by considering the expressions (50).

These contain the quantity $f\left(t - \frac{x_1}{q_1}\right)$, where, by (51):

$$x_1 = \pm i\xi\theta' + \eta \sin \theta_1.$$

If we therefore substitute the expression (53) for f, the exponent contains the real member $\pm \frac{\omega\xi\delta'}{q_1}$, and, if the solution is to be of use, this must not become ∞ for $\xi = \infty$—that is, at an infinite distance from the boundary surface. Consequently the *lower* sign is everywhere to be taken and we have:

$$\cos \theta_1 = -i\theta' \quad . \quad . \quad . \quad . \quad . \quad (58)$$

and correspondingly:

$$\mu_1 = 2 \cos \frac{\delta}{2} e^{i\frac{\delta}{2}} \quad . \quad . \quad . \quad . \quad . \quad (59)$$

Accordingly, by (50), the field-components of the refracted wave become:

$$\left.\begin{aligned}
E_\xi &= -\frac{2 \sin \theta_1}{\sqrt{\epsilon_1}} \cos \frac{\delta}{2} \cdot e^{-\frac{\omega\theta'\xi}{q_1}} \cos \left\{\omega\left(t - \frac{\eta \sin \theta_1}{q_1}\right) + \frac{\delta}{2}\right\} \\
E_\eta &= \frac{2\theta'}{\sqrt{\epsilon_1}} \cdot \cos \frac{\delta}{2} \cdot e^{-\frac{\omega\theta'\xi}{q_1}} \sin \left\{\omega\left(t - \frac{\eta \sin \theta_1}{q_1}\right) + \frac{\delta}{2}\right\} \\
E_\zeta &= 0, \ H_\xi = 0, \ H_\eta = 0
\end{aligned}\right\} (60)$$

$$H_\zeta = 2 \cos \frac{\delta}{2} e^{-\frac{\omega\theta'\xi}{q_1}} \cos \left\{\omega\left(t - \frac{\eta \sin \theta_1}{q_1}\right) + \frac{\delta}{2}\right\} \quad . \quad . \quad . \quad (61)$$

As we see, even in the case of total reflection appreciable vibrations occur in the second, optically less dense, medium. But these vibrations have the peculiarity that their phase does not depend at all on the co-ordinate ξ which is normal to the boundary plane. This causes the wave to advance in the second body in a direction parallel to the boundary plane, in the direction η, and consequently does not penetrate into the interior of the body at all. Rather, it remains confined within a boundary layer whose thickness is of the order of magnitude of a single wave-length λ_1. The greater we

assume the angle of incidence to be, the less the wave penetrates into the second body, and the more slowly it advances along the boundary plane.

For the *reflected* wave the expressions become much simpler, in that here the wave-normal:

$$x' = -\xi \cos\theta + \eta \sin\theta \quad . \quad . \quad . \quad . \quad (62)$$

which enters into the argument $t - \dfrac{x'}{q}$ of the wave-function, is real. On the other hand, by using (58) and (57) we get for the coefficient μ from (23) the complex value:

$$\mu = e^{i\delta} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (63)$$

Hence from (52) we get for the field-components of the reflected wave:

$$\left.\begin{aligned} \boldsymbol{E}_{x'} &= 0,\ \boldsymbol{E}_{y'} = \frac{1}{\sqrt{\epsilon}} \cos\left\{\omega\left(t - \frac{x'}{q}\right) + \delta\right\},\ \boldsymbol{E}_{z'} = 0 \\ \boldsymbol{H}_{x'} &= 0,\ \boldsymbol{H}_{y'} = 0,\ \boldsymbol{H}_{z'} = \cos\left\{\omega\left(t - \frac{x'}{q}\right) + \delta\right\} \end{aligned}\right\} \quad . \quad (64)$$

We can, of course, assure ourselves subsequently, by means of a simple calculation, that the expressions (55), (60), (61), (64) for the incident, refracted and reflected waves in actual fact satisfy all the conditions in the interior ($\xi \gtrless 0$) and at the boundary ($\xi = 0$) of the two bodies.

It is a characteristic of the reflected wave that it has the same amplitude as the incident wave, but in contrast with ordinary reflection, a phase which is displaced by an angle δ with respect to the incident wave.

This abrupt change of phase is equal to zero, by (57), for the limiting angle ($\sin\theta = n,\ \ \theta = 0$), where the total reflection coincides with ordinary reflection, and increases in the same sense as the angle of incidence θ increases until it assumes the value π for grazing incidence $\left(\theta = \dfrac{\pi}{2}\right)$.

The above results hold for the *f*-wave which vibrates

in the plane of incidence. It is obvious that the g-wave which vibrates perpendicularly to the incident plane obeys fully analogous laws, which nevertheless are distinguished in a characteristic way from those which have hitherto been obtained. Since the process of calculation is exactly the same it will be sufficient to summarize the results briefly :

$$\left.\begin{aligned} &\boldsymbol{E}_x = 0 && \boldsymbol{H}_x = 0 \\ &\boldsymbol{E}_y = 0 && \boldsymbol{H}_y = -\cos\omega\left(t - \frac{x}{q}\right) \\ &\boldsymbol{E}_z = \frac{1}{\sqrt{\epsilon}}\cos\omega\left(t - \frac{x}{q}\right) && \boldsymbol{H}_z = 0 \end{aligned}\right\} \quad . \quad (65)$$

and if we set :

$$\frac{\theta'\sin\theta}{\cos\theta\sin\theta_1} = \tan\frac{\tau}{2}\,(0 < \tau < \pi) \quad . \quad . \quad (66)$$

we get from (24) and (23) :

$$\sigma_1 = 2n\cos\frac{\tau}{2}\cdot e^{i\frac{\tau}{2}}, \quad \sigma = e^{i\tau} \quad . \quad . \quad . \quad (67)$$

and hence, by (50), the field components of the refracted wave become :

$$\boldsymbol{E}_\xi = 0,\ \boldsymbol{E}_\eta = 0$$

$$\left.\begin{aligned} &\boldsymbol{E}_\zeta = \frac{2}{\sqrt{\epsilon}}\cos\frac{\tau}{2}\,e^{-\frac{\omega\theta'\xi}{q_1}}\cos\left\{\omega\left(t - \frac{\eta\sin\theta_1}{q_1}\right) + \frac{\tau}{2}\right\} \\ &\boldsymbol{H}_\xi = 2\sin\theta\cos\frac{\tau}{2}\,e^{-\frac{\omega\theta'\xi}{q_1}}\cos\left\{\omega\left(t - \frac{\eta\sin\theta_1}{q_1}\right) + \frac{\tau}{2}\right\} \\ &\boldsymbol{H}_\eta = -2n\theta'\cos\frac{\tau}{2}\,e^{-\frac{\omega\theta'\xi}{q_1}}\sin\left\{\omega\left(t - \frac{\eta\sin\theta_1}{q_1}\right) + \frac{\tau}{2}\right\} \\ &\boldsymbol{H}_\zeta = 0 \end{aligned}\right\} \quad (68)$$

and by (52) the field-components of the *reflected* wave become :

$$\left.\begin{aligned} &\boldsymbol{E}_{x'} = 0,\ \boldsymbol{E}_{y'} = 0,\ \boldsymbol{E}_{z'} = \frac{1}{\sqrt{\epsilon}}\cos\left\{\omega\left(t - \frac{x'}{q}\right) + \tau\right\} \\ &\boldsymbol{H}_{x'} = 0,\ \boldsymbol{H}_{y'} = -\cos\left\{\omega\left(t - \frac{x'}{q}\right) + \tau\right\},\ \boldsymbol{H}_{z'} = 0 \end{aligned}\right\} \quad . \quad (69)$$

The wave g which vibrates perpendicularly to the incident plane thus experiences a different phase change τ at total reflection from the wave f which vibrates in the incident plane; by (57) and (66) we have in general that $\delta > \tau$. It is only in the extreme cases of the limiting angle $\left(\theta_1 = \frac{\pi}{2}\right)$ and of grazing incidence $\left(\theta = \frac{\pi}{2}\right)$ that δ and τ coincide.

§ 14. To solve the problem of total reflection entirely we have yet to generalize the particular solution that has been found for the case of any arbitrarily given waves $f(t)$ and $g(t)$. This can be done simply by first multiplying the expression (53) by means of any arbitrary complex constant, by which the vibration acquires an arbitrary amplitude and an arbitrary phase-constant, and then performing a summation over the whole multiples of a definite radian frequency ω. This causes the real part of $f(t)$ or $g(t)$ respectively to assume the form of a Fourier series (II, § 38):

$$\sum_{n=1}^{\infty} C_n \cos\left(\frac{2\pi n t}{\boldsymbol{T}} + \theta_n\right) \quad . \quad . \quad . \quad . \quad (70)$$

and it is known that *every* function which is periodic with respect to the time $\boldsymbol{T}$ can be represented by such a series. By (12) its initial term C_0 is set equal to zero. In assuming the wave-functions f and g to be periodic we are not actually introducing a limitation, for we can make the period $\boldsymbol{T}$ so great that for the measurement only those times t come into question which lie within a period—that is, between 0 and $\boldsymbol{T}$. This occurs, for example, if we denote the moment of time at which the source begins to emit light by $t = 0$, and the moment of time at which the last measurement has been performed by $t = \boldsymbol{T}$.

CHAPTER II

SPECTRAL RESOLUTION. INTERFERENCE. POLARIZATION

§ 15. In deriving the laws of reflection and refraction in the preceding chapter no assumptions were made about the form of the wave-functions f and g which were used. This also applies to the laws of total reflection. For the expansion of a wave-function into a Fourier series is only a particular form of mathematical representation. In particular we must guard ourselves against thinking that the wave-functions must have any properties of periodicity within the large fundamental period $\boldsymbol{T}$. Rather they can even have the value zero at times, and can behave quite arbitrarily before and after these zero values. Hence it also follows that no objective meaning for the form of the wave-function can be ascribed to the amplitudes C_n of the Fourier series (70), nor to the fundamental period $\boldsymbol{T}$ on which they are dependent.

To get information about the form of a wave it is therefore necessary to make special measurements. The method which suggests itself most readily in principle is that which was used by H. Hertz to find the form of the electromagnetic waves which he had discovered—namely, to produce *stationary vibrations* by means of reflection from a perfectly conducting and hence perfectly reflecting surface (III, § 92). For it is in this process that the periodic character of a wave manifests itself, owing to the appearance of equidistant nodes and anti-nodes, which allow the wave-length, and hence the frequency, to be measured. The corresponding experiment with optical waves was first successfully performed by O. Wiener (1890), who allowed the violet rays of an arc lamp to fall

normally on a silver mirror. To make the resulting nodes and anti-nodes of the electric field-strength visible, a very thin membrane-like layer of collodion containing silver chloride, which is sensitive to light and does not appreciably disturb the vibrations of light, was placed obliquely at a small angle α over the mirror (Fig. 2). No photographic action was then found to occur at the places where the membrane encountered a nodal plane of the electric field-strength, for example at the mirror itself, whereas a maximum darkening of the membrane occurred at the places midway between the nodes, where the planes of the anti-nodes (shown by the broken lines in the figure) are intersected by the membrane. The wave-length can then be directly calculated from the distance between the dark stripes and the angle of inclination of the membrane to the mirror.

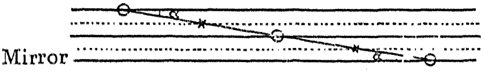

FIG. 2.

On account of the smallness of the wave-lengths of ordinary light, it is customary to use as the unit of length in optics the micron, $1\mu = 10^{-4}$ cm., or $1\mu\mu$ (millimicron) $= 10^{-7}$ cm., or, more often, the Ångström unit, $1\text{Å} = 10^{-8}$ cm. The visible spectrum stretches from about $0{\cdot}4\mu$ (violet) to about $0{\cdot}8\mu$ (red).

§ 16. Although this experiment proves the periodic character of a light-wave, it by no means follows from this that the Fourier series (70) reduces for this wave to a single term, a single order number n. For the measurement, even in the case of the sharpest spectral lines, always admits of a very great range of order numbers, whose frequencies vary but imperceptibly. For since:

$$\omega = \frac{2\pi n}{T} = 2\pi\nu \quad . \quad . \quad . \quad . \quad . \quad (71)$$

and since T is enormously great compared with the time occupied by a single light vibration, the order numbers n

that come into question are enormously great and a change in the order number n by one or more units effects no appreciable change in ω.

Thus we must conceive of light of a definite colour, so-called monochromatic or "homogeneous" light, not as a wave of a single period, like the sound-wave of a definite tone in acoustics, but as composed of numerous waves of almost equal periods. There is no monochromatic light in the absolute sense, but only in a more or less approximate sense. Expressed mathematically, the condition for homogeneous light of frequency $\omega_0 = \frac{2\pi n_0}{T}$ is that in the Fourier series (70) only those amplitudes C_n differ from zero for which :

$$\frac{n - n_0}{n_0} << 1 \quad . \quad . \quad . \quad . \quad . \quad (72)$$

The degree of homogeneity is determined by the greatest value which the ratio (72) can assume without C_n vanishing.

A particularly clear picture of the behaviour of the wave-function of homogeneous light is obtained by substituting in the Fourier series (70) :

$$n = n_0 + (n - n_0)$$

and expanding the cosine correspondingly. The series (70) can then be written as the single term :

$$C_0 \, . \cos\left(\frac{2\pi n_0 t}{T} + \theta_0\right) \quad . \quad . \quad . \quad . \quad (73)$$

where :

$$C_0 \cos \theta_0 = \Sigma C_n \cos\left(\frac{2\pi(n - n_0)t}{T} + \theta_n\right)$$

$$C_0 \sin \theta_0 = \Sigma C_n \sin\left(\frac{2\pi(n - n_0)t}{T} + \theta_n\right)$$

and can be regarded as a single vibration of frequency ω_0, amplitude C_0 and phase-constant θ_0. It is true that C_0 and θ_0 are not strictly constant, but on account of (72) they only change relatively slowly, and in general

irregularly, with the time t; in fact, the more slowly the more homogeneous the light.

This shows the close relationship which exists in the case of homogeneous light between the fluctuations of the amplitude and the phase on the one hand, and the degree of homogeneity on the other. An absolutely constant amplitude and an exactly regular phase could be possible only in the case of absolutely homogeneous light; every kind of fluctuation denotes a lack of homogeneity.

§ 17. A far sharper resolution than that given by reflection at an opaque mirror is obtained by the reflection or transmission of light through a transparent *plane-parallel plate*. This problem, too, may in general be solved by means of the methods here developed.

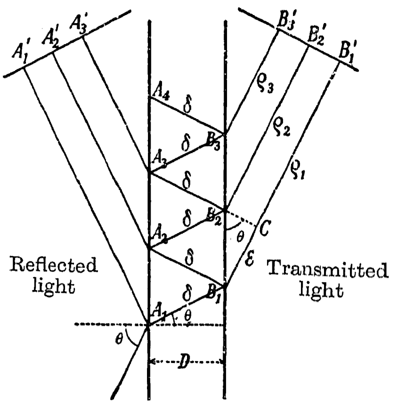

FIG. 3.

If a plane wave falls on a transparent plane plate of thickness D at an angle of incidence θ (Fig. 3), it is partly reflected, it partly penetrates into the plate; the penetrated portion of the wave partly travels to and fro within the plate between its boundary faces and partly passes out through the rear or the front face of the plate.

The complete solution of the problem is contained in the following assumption. In the first medium (air) there are two waves, one directed towards the plate at the angle of incidence θ, the other directed away from the plate at the angle of reflection θ. In the second medium (glass) there are also two waves, both making an angle θ_1 with the normal to the plate, the one directed from the front to the rear boundary surface, the other moving in the contrary direction. In the third medium (air) there is a

single wave which advances in a direction parallel to the original incident wave.

These five waves are connected by four boundary conditions—namely, one each for the electric field-strength and the magnetic field-strength at each of the two parallel boundary surfaces. From these it is possible uniquely to determine the other four waves, when the incident wave is given.

We achieve the same object, not so directly but more concretely than by the preceding method, if we take as our starting-point the course of the process in time; this has the fundamental advantage that it can then also be applied when, as always happens in reality, the cross-section of the incident wave and the surface of the plate does not exceed all limits.

Let the incident wave $f\left(t - \frac{x}{q}\right)$ be represented graphically in Fig. 3 by any arbitrarily selected ray, which we assume strikes the first surface of the plate at A_1, so that x represents the wave-normal at the point A_1. The reflected ray (that is, the reflected wave) then passes from A_1 out into the air again with the coefficient of absorption or "weakening" (*Schwächungskoeffizient*) μ, which is given by (23); whereas the refracted ray, which has the coefficient of absorption μ_1, gets as far as B_1 in the plate. Here it splits up again, partly to enter the air in the forward direction and partly to turn back in the plate towards A_2. And so the same process continues until finally the energy of the ray is essentially used up and the remainder can be neglected. According to this method of reasoning, the four waves which were considered above and which are to be calculated from the incident wave present themselves as four sums having an infinite number of terms each of which is known and can be specified.

Let us first consider the *transmitted* wave by choosing any wave-plane B'_1, B'_2, B'_3 . . . sufficiently far from the plate. This wave arises from the superposition of those waves which are represented graphically by the rays

$A_1B_1B'_1$, $A_1B_1A_2B_2B'_2$, ... If we denote the length $A_1B_1 = B_1A_2 = A_2B_2 =$ . . . by δ, and the lengths $B_1B'_1$, $B_2B'_2$, . . . by ρ_1, ρ_2, . . . then the expressions for the wave-functions which meet in the selected wave-plane and which correspond to the individual rays result if we substitute the light-paths quoted in the argument of the function and add, besides, the absorption coefficients μ, μ_1, μ', μ'_1 which present themselves at every reflection and refraction. Here μ' and μ'_1 denote the coefficients which are valid for a wave which passes from glass into air. They are derived from the expressions for μ and μ_1 by exchanging the angles θ and θ_1 with each other.

Accordingly the wave-function of the first ray, which has undergone two refractions, at A_1 and at B_1, runs, at the point B'_1:

$$\mu_1\mu'_1 \cdot f\left(t - \frac{x}{q} - \frac{\delta}{q_1} - \frac{\rho_1}{q}\right)$$

and the wave-function of the second ray, which has undergone two refractions at A_1 and at B_2 and two reflections at B_1 and A_2, is, at the point B'_2:

$$\mu_1\mu'^2\mu'_1 \cdot f\left(t - \frac{x}{q} - \frac{3\delta}{q_1} - \frac{\rho_2}{q}\right), \text{ and so forth.}$$

Hence we get for the wave-function of the whole of the transmitted light:

$$\sum_{p=0}^{\infty} \mu_1(\mu')^{2p}\mu'_1 f\left(t - \frac{x}{q} - \frac{(2p+1)\delta}{q_1} - \frac{\rho_{p+1}}{q}\right) \quad . \quad (74)$$

Here the coefficients μ_1, μ' and μ'_1 can be reduced in a simple manner, in view of (23) and (24), to terms of the single coefficient μ by means of the equations:

$$\mu' = -\mu, \ \mu_1 = 1 + \mu, \ \mu'_1 = 1 - \mu \quad . \quad . \quad (75)$$

Concerning the argument of f, we can, by (14), express the traversed lengths in terms of a common denominator:

$$-\frac{x + (2p+1)\delta n + \rho_{p+1}}{q} \quad . \quad . \quad . \quad (76)$$

and so reduce the total path of the light to the equivalent path in air (" optical length of path "). We can then reduce the lengths ρ_{p+1} to terms of the first of these, ρ_1, by means of the following relation which is immediately evident from Fig. 3 :

$$\left.\begin{aligned} \rho_1 - \rho_2 &= B_1C = \epsilon \\ \rho_1 - \rho_3 &= 2\epsilon, \ldots \\ \rho_1 - \rho_{p+1} &= p\epsilon \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (77)$$

Finally we have the following expressions, which can be easily obtained for the lengths δ and ϵ which have been introduced :

$$\delta = \frac{D}{\cos\theta_1} \quad . \quad . \quad . \quad . \quad . \quad . \quad (78)$$

$$\left.\begin{aligned} \epsilon &= B_1B_2 \,.\, \sin\theta = 2\delta\sin\theta_1\sin\theta \\ \epsilon &= 2D\tan\theta_1\sin\theta \end{aligned}\right\} \quad . \quad . \quad (79)$$

If we introduce the following abbreviations for the terms of the argument f not involving p :

$$t - \frac{x + \delta n + \rho_1}{q} = \alpha \quad . \quad . \quad . \quad . \quad . \quad (80)$$

and for those multiplied by p :

$$\frac{2\delta n - \epsilon}{q} \cdot p = \beta p \,. \quad . \quad . \quad . \quad . \quad . \quad (81)$$

then the wave-function (74) now runs :

$$\sum_{p=0}^{\infty} (1 - \mu^2) \,.\, \mu^{2p} \,.\, f(\alpha - p\beta) \quad . \quad . \quad . \quad . \quad (82)$$

§ 18. In order to be able to perform this summation we require to make some assumption about the form of f. We therefore suppose f to be expanded into a Fourier series and first consider a single term of this series of frequency ω. Since it is often more convenient to calculate with exponential functions than with trigonometrical functions, we regard the real vibration as the real part of a complex vibration and so write :

$$f(t) = e^{i\omega t} \,. \quad . \quad . \quad . \quad . \quad . \quad . \quad (83)$$

where, for the sake of brevity, we follow the usual custom and suppress the symbol $\boldsymbol{R}$.

The expression (82) then becomes:

$$\sum_{p=0}^{\infty}(1-\mu^2)\mu^{2p}\,e^{i\omega(\alpha-p\beta)}$$

and this sum is equal to:

$$\frac{(1-\mu^2)\,e^{i\omega\alpha}}{1-\mu^2\,e^{-i\omega\beta}} \quad . \quad . \quad . \quad . \quad . \quad (84)$$

The real part of this complex expression represents the required wave of the transmitted light. Since the variables t and ρ_1 are contained only in α, it is simply periodic and has the radian frequency ω. By (10) its intensity is proportional to the square of the amplitude. This square can be most simply calculated by observing that it is at the same time the square of the absolute value of the complex quantity (84); that is, it is the product of (84) and its conjugate imaginary. Since we are concerned only with the ratio of the intensity of the transmitted light to the intensity of the incident light, and since the absolute value of (83) is equal to 1, we obtain for the ratio of the intensity of the transmitted light to that of the incident light:

$$J=\frac{(1-\mu^2)\,.\,e^{i\omega\alpha}}{1-\mu^2\,e^{-i\omega\beta}}\cdot\frac{(1-\mu^2)\,.\,e^{-i\omega\alpha}}{1-\mu^2\,e^{i\omega\beta}}$$

$$J=\frac{(1-\mu^2)^2}{1-2\mu^2\cos\omega\beta+\mu^4} \quad . \quad . \quad . \quad . \quad (85)$$

By (81), (78), (79) and (20) we here have:

$$\beta=\frac{2nD\cos\theta_1}{q}=\frac{2D}{q_1}\cos\theta_1$$

Thus:

$$\omega\beta=\frac{4\pi D}{\lambda_1}\cos\theta_1=\xi \quad . \quad . \quad . \quad . \quad (86)$$

where λ_1 denotes the wave-length in the substance of the plate, and:

$$J=\frac{(1-\mu^2)^2}{1-2\mu^2\cos\xi+\mu^4} \quad . \quad . \quad . \quad . \quad (87)$$

Maximum intensity is attained when $\xi = 2\pi \, . \, p$, that is, when :

$$\frac{2D}{\lambda_1} \cos \theta_1 = p \text{ (integer)} \quad . \quad . \quad . \quad . \qquad (88)$$

and then $J = 1$. There are intermediate minima of intensity, of amount :

$$J = \left(\frac{1 - \mu^2}{1 + \mu^2}\right)^2 \quad . \quad . \quad . \quad . \quad . \qquad (89)$$

The differences between the maxima and minima become the greater the more μ approximates to 1—that is, the greater the angle of incidence θ.

But what characterizes the phenomenon here considered most sharply in its bearing on the measurements is the steepness of these maxima. This is due to the fact that the parameter ξ in (86), and consequently also p in (88), is in general a very great number, on account of the difference in the orders of magnitude of D and λ_1. Consequently the intensity of the light varies very markedly with the ray direction θ.

Now we must bear in mind, as already emerges from the considerations of § 5, and as will be further explained in § 36, that in optics we never observe a single definite ray direction or a single system of parallel wave-planes, but always only a cone of ray directions, which may be only very narrow. Thus if we ascribe to every ray direction θ the point of an infinitely distant screen at which all rays parallel to this direction aim, then there will always be a great number of such adjacent points θ on which the light will impinge, and since J varies greatly with θ, a great number of maxima and minima of intensity will appear next to each other on the screen, the distances between them being regulated by the successive order numbers p in (88). Now if μ is nearly equal to 1, the intensity J, by (87), is in general weak, and it is only when the condition (88) is exactly fulfilled that it becomes equal to 1—that is, the maxima stand out in sharp contrast on the dark background.

All these considerations refer to a perfectly definite frequency ω of the wave—that is, to absolutely homogeneous light. Since, as we have seen, real waves always contain many, although very adjacent, frequencies, the observed radiation always contains several systems of maxima and minima of the type considered, each system having the maxima and minima at different intervals; and in every ray of the transmitted light or, respectively, at the point θ on the infinitely distant screen several rays of different colours meet, and in general they are of different intensity. It may also happen that two different colours have their maximum intensity at a definite point θ, so that by (88):

$$p\lambda_1 = p'\lambda'_1$$

or, referred to air:

$$p\lambda = p'\lambda' \quad . \quad . \quad . \quad . \quad . \quad . \quad (90)$$

Then the maximum of order p for the wave-length λ coincides with the maximum of order p' for the wave-length λ'. If λ and λ' differ only slightly the nearest maxima of the two wave-lengths, whose order numbers p and p' differ by only a few units, will fall close together, and the two systems of maxima and minima will also nearly coincide in the immediate neighbourhood. But if we proceed to higher order numbers the little difference in the distances between the maximum and minimum will make itself felt, and the maxima of the two systems will move apart, so that the maxima of the greater wave-length, say λ', will, on account of their greater separation, move ahead of those of the smaller wave-length λ of the corresponding order number. Finally, when the order number is sufficiently increased, the maxima of λ' with the order number $p' + r$ will, so to speak, catch up those of λ with the next higher order number $p + r + 1$. Then the two systems will again coincide, but in such a way that the order number of the maximum of λ will have increased by one more than the maximum of λ', that is:

$$(p + r + 1)\lambda = (p' + r)\lambda'$$

This, combined with (90) gives :

$$\frac{\lambda' - \lambda}{\lambda} = \frac{1}{r} \quad . \quad . \quad . \quad . \quad . \quad . \quad (91)$$

Thus by counting up the maxima between two successive coincidences of the two systems we have a very exact method for measuring small relative differences of wave-length.

The intensity J' can be found for the *reflected* light just as has been done for the light transmitted from the plate. But it is unnecessary to carry out the calculation in detail. For the result comes out at once by applying the energy principle, which states that the sum of the intensity J' of the reflected light and of the intensity J of the transmitted light is equal to the intensity 1 of the incident light. Thus by (87) :

$$J' = 1 - J = \frac{4\mu^2 \sin^2 \frac{\xi}{2}}{1 - 2\mu^2 \cos \xi + \mu^4} \quad . \quad . \quad (92)$$

The reflected light is, as we say, "complementary" to the transmitted light. The minima of intensity are zero and stand out sharply from the surroundings when μ is nearly equal to 1. In other respects, the same laws hold here as for transmitted light.

Hitherto we have considered only the light which vibrates in the incident plane—namely, the f-wave. But it is evident that the same inferences may be drawn also for the light which vibrates perpendicularly to the incident plane, namely the g-wave. The only difference in the results is that in the case of the g-wave the coefficient σ (§ 9) takes the place of the coefficient μ. The position of the maxima and minima remains precisely the same as in the case of the f-wave.

§ 19. According to the equation (86) the angle ξ which is characteristic of the intensity of the light depends not only on the ray-direction θ, but also on the thickness of plate D. This circumstance is made use of in interferometers of variable thickness (air plate) in order to measure

the dependence of the intensity of light J on the thickness of layer D for the case where the light is incident normally ($\theta_1 = 0$). Then by (88) those thicknesses of layer for which the transmitted light has the maximum intensity 1 are whole multiples of the half wave-length in the substance of the layer :

$$D = p \cdot \frac{\lambda_1}{2} \quad \dots \dots \quad (93)$$

This is easy to understand if we reflect that when the thickness of layer is a multiple of the half wave-length the phases of the waves which are superposed on each other after several reflections within the layer before passing through it differ by 2π, so that all the waves reinforce each other.

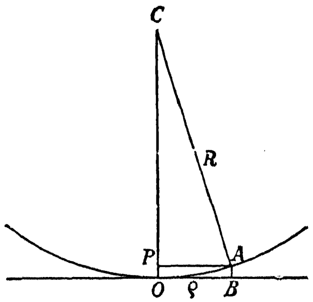

FIG. 4.

A phenomenon which exhibits the effect of several thicknesses of layer simultaneously is that of *Newton's Rings*. If the curved surface of a plano-convex glass lens is placed in close contact on the surface of a plane-parallel glass plate (Fig. 4) a thin air-film is formed, whose thickness increases from the value zero to the value D at the distance ρ from O. If we now allow homogeneous light to fall in a parallel beam normally on to the plate, we see in the transmitted beam a system of light and dark rings with a bright spot in the centre, and in the reflected light we see the complementary pattern.

An approximate theory of this phenomenon is obtained very simply from our preceding remarks. If we restrict ourselves to values of ρ which are small compared with the radius R of the sphere, we may write :

$$D = AB = PO = R - \sqrt{R^2 - \rho^2}$$

or :

$$D = \frac{1}{2}\frac{\rho^2}{R}$$

Consequently, by (93), we obtain for the maxima of intensity in the transmitted light or, respectively, for the minima in the reflected light, if λ denotes the wave-length in air :

$$\rho = \sqrt{p\lambda R} \quad . \quad . \quad . \quad . \quad . \quad (94)$$

Thus the radii of the dark rings in the reflected light are in the ratio of the square roots of the series of natural numbers.

For two colours which differ only slightly from each other, such as the two components of the yellow line of sodium (5890 Å and 5896 Å) the ring systems differ slightly and move apart more and more as the distance from the central point O increases, so that they blur each other's effects until for, say, $p = r$ the maxima and minima, respectively, coincide again, as can be seen from the renewed sharpness of the again common ring-system. This allows us to calculate, by (91), the relative difference of wave-length of the two sodium lines.

§ 20. In the preceding section we have been led to consider the combined action of several trains of waves superposed on each other and moving in the same direction; we shall now treat the same question of the superposition of light waves systematically for the general case.

Let $f_1\left(t - \frac{x}{q}\right)$ and $f_2\left(t - \frac{x}{q}\right)$ be the wave-functions of two rays which vibrate in the same direction and which propagate themselves in the same direction. They combine to form a single ray whose wave-function is :

$$f = f_1 + f_2 \quad . \quad . \quad . \quad . \quad . \quad (95)$$

It follows from this that the intensity J of the resulting rays depends on the intensities J_1 and J_2 of the two com-

ponents in the following way, if we disregard a factor which is of no importance :

$$J = \overline{f^2} = \overline{f_1^2} + \overline{f_2^2} + 2\overline{f_1 f_2}$$
$$J = J_1 + J_2 + 2\overline{f_1 f_2} \quad . \quad . \quad . \quad . \quad . \quad (96)$$

Thus the intensity of the composite ray deviates in general from the sum of the intensities of the individual rays. This phenomenon is called the "interference" of the two rays 1 and 2. Under certain circumstances no interference occurs—namely, when the wave-functions f_1 and f_2 are fully independent of one another—that is, when the rays 1 and 2 are "non-coherent." For then, as in (43), we have :

$$\overline{f_1 f_2} = \overline{f_1} \cdot \overline{f_2} = 0.$$

This is always the case, for example, when the two rays come from different sources. Hence two such rays, even if they have exactly the same colour, never interfere with each other. But non-coherence is not a necessary condition for the absence of interference. Rays can even be "fully coherent" and yet not interfere: we take "fully coherent" to mean that one of the two wave-functions is completely determined by the other. This occurs, for example, in the case of homogeneous light, if the arguments of f_1 and f_2 always differ by an odd multiple of $\frac{\pi}{2}$. Then, in spite of complete coherence, $\overline{f_1 f_2} = 0$. The general case is exemplified by "partial coherence," where f_1 and f_2 are not independent of each other, and yet are not completely defined by each other.

§ 21. A somewhat different way of regarding the same problem results if we write the two wave-functions f_1 and f_2 as Fourier series having the same fundamental period T. Then the resulting wave-function $f = f_1 + f_2$ may also be represented directly as a Fourier series by compounding each two partial waves of the two series, which have the same frequency, into a single partial

wave. We then have, for example, for the radian frequency ω of the partial waves :

$$\left.\begin{array}{l} a_1 \cos\left\{\omega\left(t - \frac{x}{q}\right) + \theta_1\right\} \\ \text{and:} \quad a_2 \cos\left\{\omega\left(t - \frac{x}{q}\right) + \theta_2\right\} \end{array}\right\} \quad . \quad . \quad . \quad (97)$$

In summing these two expressions an important part is played by the phase-difference of the two waves :

$$\theta_1 - \theta_2 = \Delta. \quad . \quad . \quad . \quad . \quad . \quad (98)$$

Instead of the phase-angle θ, sometimes the constant length d or the time constant δ is introduced into the wave-function by writing :

$$a \cos\left\{2\pi\left(\frac{t}{\tau} - \frac{x + d}{\lambda}\right)\right\}$$

or :

$$a \cos\left\{2\pi\left(\frac{t + \delta}{\tau} - \frac{x}{\lambda}\right)\right\}$$

and we then speak of the "difference of path," $d_1 - d_2$, of the two waves or of the "retardation," $\delta_1 - \delta_2$, of the second wave with respect to the first. The phase difference 2π corresponds to a difference of path of one wave-length, λ, or a retardation of one period of vibration, τ. By adding the two expressions (97) we get for the resulting partial wave :

$$a \cos\left\{\omega\left(t - \frac{x}{q}\right) + \theta\right\} \quad . \quad . \quad . \quad . \quad (99)$$

where :

$$\left.\begin{array}{l} a \cos\theta = a_1 \cos\theta_1 + a_2 \cos\theta_2 \\ a \sin\theta = a_1 \sin\theta_1 + a_2 \sin\theta_2 \end{array}\right\} \quad . \quad . \quad (100)$$

Thus the amplitude and the phase-angle of the resultant wave are formed from those of its components according to the same rules as the absolute value and the direction of a vector which results from two other vectors.

From (100) and (98) we obtain for the square of the amplitude a :

$$a^2 = a_1{}^2 + a_2{}^2 + 2a_1a_2 \cos\Delta \quad . \quad . \quad . \quad (101)$$

This equation expresses that two absolutely homogeneous waves of the same frequency moving in the same direction and with their vibrations in the same direction interfere in general. This does not contradict the law of § 20 that two rays coming from different sources of light and having the same colour never interfere. For, according to § 16, there is no absolutely homogeneous light in nature; rather, even the most near homogeneous optical ray always contains very many partial waves of nearly the same vibration frequency; each of these partial waves interferes with the partial waves of the same frequency in the other ray. Now if the two rays are non-coherent the phase-difference Δ changes in a quite irregular manner in passing from one pair of waves to another, so that no appreciable interference effect can occur.

§ 22. We shall next consider the composition of two waves which propagate themselves in the same direction, but whose directions of vibration are at right angles to each other, as in § 11 except that now we start from the Fourier expansion. As in § 11 let z be the direction of propagation, x the direction of the electric field-strength, y that of the second wave. We can then represent the electric field-strengths of two partial waves of the same frequency ω after the model of (97) by:

$$\left.\begin{aligned} a\cos\left\{\omega\left(t-\frac{z}{q}\right)+\theta_1\right\} &= x \\ b\cos\left\{\omega\left(t-\frac{z}{q}\right)+\theta_2\right\} &= y \end{aligned}\right\} \quad . \quad . \quad (102)$$

where we choose appropriate units for x and y. Without affecting the generality of our problem we may take $a \geqq b \geqq 0$. This method of representation enables us to form a very clear picture of the law, according to which the resultant electric field-strength of a definite wave-plane z varies in magnitude and direction with the time, since it is represented at every moment by the position

of the point x, y which moves about in the plane $z =$ const.

As for the orbit of this point, it is obtained by eliminating the time t from the equations (102), which is done most simply by calculating the values of $\cos \omega\left(t - \frac{z}{q}\right)$ and $\sin \omega\left(t - \frac{z}{q}\right)$, then squaring and adding them and putting their sum equal to 1. The equation of the path then comes out as:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{2xy}{ab}\cos\Delta = \sin^2\Delta. \quad . \quad . \quad (103)$$

This is an ellipse which degenerates into a straight line when $\Delta = n\pi$; whereas when $\Delta = (n + \frac{1}{2})\pi$ its axes coincide with the co-ordinate axes. In general the angle ϕ, which one axis of the ellipse forms with the x-axis, is determined by the relation:

$$\tan 2\phi = \frac{2ab\cos\Delta}{a^2 - b^2} \quad . \quad . \quad . \quad . \quad (104)$$

which can be obtained directly from (45) if we insert the following values, which correspond to our present case:

$$A = \overline{f^2} = \frac{a^2}{2}, \; B = \overline{g^2} = \frac{b^2}{2},$$

$$C = \overline{fg} = \frac{ab}{2}\cos\Delta.$$

If, further, we inquire into the velocity with which the elliptic orbit is traversed, we get from (102) by differentiation:

$$xdy - ydx = \omega ab \sin\Delta \, . \, dt \quad . \quad . \quad . \quad (105)$$

This signifies that the motion of the reference-point x, y occurs in accordance with the law of sectional areas (I, § 50), and moreover, since $ab > 0$, it occurs in the positive or negative sense according to the sign of $\sin\Delta$. Since the z-axis of our right-handed co-ordinate system (I, § 16) is the direction of propagation of the light, the

sense in which the rotation represented by the motion occurs has a definite physical meaning.

In optics light is called dextrorotatory (or dextrogyrous) if the motion of the reference-point x, y is clockwise to an observer looking towards the light. For this reason we have to regard the sense of rotation of dextrorotatory light as negative, and that of lævorotatory light as positive.

We can obtain a clear conception of the way in which the motion of the reference-point depends on the constants a, b, Δ by keeping, say, a and b constant, but allowing the phase-difference Δ to increase from 0 to 2π. When $\Delta = 0$ the elliptic motion degenerates into a rectilinear vibration, which by (104) or (102) is inclined to the x-axis and lies in the first and third quadrants, the angle of inclination being:

$$\phi = \tan^{-1}\frac{b}{a} \leqq \frac{\pi}{4} \quad . \quad . \quad . \quad . \quad (105a)$$

and the amplitude $\sqrt{a^2 + b^2}$.

Now if the phase-difference Δ increases, the ellipse widens and is traversed in the positive sense. The light is then lævorotatory. At the same time the angle ϕ becomes smaller—that is, the major axis of the ellipse moves towards the x-axis. When $\Delta = \frac{\pi}{2}$ the ellipse attains its greatest curvature, the semi-axes are a and b, and their directions coincide with those of the co-ordinate axes. If Δ increases beyond $\frac{\pi}{2}$, the major axis passes over into the second and fourth quadrants and the ellipse at the same time becomes flatter and flatter until, when $\Delta = \pi$, it again shrinks together to a straight line which is inclined at the angle given by (105a) to the x-axis. From then on, when Δ increases still further, the form of the ellipse is exactly repeated in the reverse direction, as we see from the fact that the orbital equation (103) remains unchanged if Δ is replaced by $2\pi - \Delta$. The

ellipse again widens and its major axis again moves towards the x-axis, which it reaches when $\Delta = \frac{3\pi}{2}$. But the orbit is now being continually traversed in the negative direction. Finally, when $\Delta = 2\pi$, the original motion is again resumed. An equally instructive example is that obtained by considering the phase-difference Δ fixed and the amplitudes a and b of the components of vibration to change gradually. Among other things we see from (104) that then the direction ϕ of the axes depends only on the ratio of the amplitudes a and b, and that the same holds for the ratio of the lengths of the axes.

If we enquire into the condition which makes the vibration circular, we obtain the answer most simply from the expression (104) for the direction of the axes, which must become indeterminate in the case of a circle —that is, must assume the form $\frac{0}{0}$. This gives us the following two conditions for a circular vibration :

$$a = b, \ \Delta = \left(n + \frac{1}{2}\right)\pi \quad . \quad . \quad . \quad (106)$$

which must be fulfilled simultaneously. The circle is traversed in the positive or negative sense according as n is even or odd.

§ 23. The elliptic vibration considered in the preceding section at the same time represents the most general case of the composition of any arbitrary number of plane waves moving in the z-direction and having the same radian frequency ω. For every linear vibration may be resolved into two linear components which vibrate in the x- and the y-direction and have the phase-difference zero, and all the vibrations which occur in the x- or the y-direction yield on being superposed, by § 21, a single vibration in the direction in question.

We can also picture the general case of the elliptic vibration in another manner besides that involving two linear vibrations—namely, by means of two circular vibrations with opposite senses of rotation. This is

obvious, if we reflect that every linear vibration may be regarded as the resultant of two circular vibrations having the same radius but being performed in opposite directions. For the two revolving points then always meet at the same points, and these points define the direction and half the amplitude of the resultant linear vibration.

In this way we obtain for the general case your circular vibrations; of these there are two pairs of vibrations which occur in the same direction and so compound in each case into a single circular vibration in the same direction: the two single resultants vibrating in opposite directions can then be compounded. The sense of the final resultant circular vibration is, of course, that of the single resultant which has the greater radius.

§ 24. Let us finally glance at the way in which the electric field-strengths x and y in (102) depend on the wave-normal z. If t is constant, these equations represent the so-called "wave-line," in the present case an elliptical helix: and by (105), if we replace the dt in it by $-\frac{dz}{q}$, we get a right-handed screw (II, § 32) if $\sin\Delta < 0$—that is, if the sense of rotation is negative. The pitch of the screw is then equal to the wave-length λ.

If we take into consideration what was said in § 23 about the sense of traverse of the dextrorotatory light, it follows that dextrorotatory light is represented by a right-handed screw, and lævorotatory light by a left-handed screw. The motion of the reference-point x, y in the plane z is obtained if we displace the helix in the direction z with the velocity q and fix our attention on its instantaneous point of intersection with the plane z = const. If we reverse the direction of displacement of the helix, a wave results which propagates itself in the negative direction of the z-axis and at the same time turns in the reverse sense, so that a right-handed screw again represents dextrorotatory light and a left-handed screw lævorotatory light.

§ 25. We shall next enquire what definite phase-differences Δ can be realized between the two vibration components x and y. If we start with a linearly polarized wave having its direction of vibration in the first and third quadrants, then in (102) the ratio of the amplitudes is constant :

$$\frac{b}{a} = \tan\phi \leqq 1 \quad . \quad . \quad . \quad . \quad . \quad . \qquad (107)$$

just as in (105a); and the phase-difference is $\theta_1 - \theta_2 = \Delta = 0$. We shall cause this wave to be reflected at some arbitrary angle of incidence, so that the electric field-strength x vibrates in the plane of incidence. For ordinary reflection ($\sin\theta < n$) no abrupt change of phase occurs, so the phase-difference Δ remains equal to zero, and the two reflected waves again combine to form a linearly polarized wave, except that, by (105a), the direction of vibration becomes changed owing to the difference in the reflection coefficients μ and σ. This has already been remarked in § 11.

But it is quite different in the case where the wave is totally reflected ($\sin\theta > n$). For here, conversely, the amplitudes a and b of the two wave-components remain unchanged, whereas a phase-difference $\Delta = \delta - \tau$ now introduces itself, the value of which is obtained from (57) and (66) ·

$$\tan\frac{\Delta}{2} = \tan\frac{\delta - \tau}{2} = \frac{\theta' \,.\, \cos\theta}{\sin\theta \,.\, \sin\theta_1} \quad . \quad . \qquad (108)$$

and θ' is determined by (56).

This phase-difference Δ is, as we have already found at the end of § 13, always positive and vanishes at the two extreme cases of the limiting angle—namely, when $\theta = 0$ and when $\theta = \frac{\pi}{2}$ (grazing incidence). Thus between these values it attains a maximum. This occurs, by (108), at the angle of incidence :

$$\sin^2\theta = \frac{2n^2}{1 + n^2} \quad . \quad . \quad . \quad . \quad . \qquad (109)$$

and the maximum value is :

$$\Delta = \pi - 4\tan^{-1}n \quad . \quad . \quad . \quad . \qquad (110)$$

$$\left(0 < \tan^{-1}n < \frac{\pi}{4}\right)$$

The smaller the index of refraction n, the greater the phase-difference Δ that can be obtained. After the reflection the two vibration components combine into an elliptic vibration, the direction of whose axes, ϕ', is by (104) and (107), given by :

$$\tan 2\phi' = \tan 2\phi \cos \Delta \quad . \quad . \quad . \qquad (111)$$

To obtain circularly vibrating light we must, by (106), produce a phase-difference of $\frac{\pi}{2}$. If we enquire what is the least value that n can have in order that the phase difference $\frac{\pi}{2}$ may be obtained, we find that it follows from (110) that :

$$n = \tan\frac{\pi}{8} = \sqrt{2} - 1 \, . \quad . \quad . \quad . \qquad (112)$$

The reciprocal of this value is $\sqrt{2} + 1 = 2{\cdot}414$. A substance must therefore have an index of refraction of at least this value if it is to be possible for the phase-difference $\frac{\pi}{2}$ to be produced in it between the f- and the g-waves by total reflection. Of the known substances only diamond fulfils this condition.

Now if a phase-difference of $\frac{\pi}{2}$ cannot be practically obtained by means of a single reflection, we can achieve this fairly conveniently by means of two successive reflections, as by a *Fresnel rhomb*. This is an oblique-angled parallelepiped of transparent glass (Fig. 5) whose angle is chosen so that a ray of light which passes through the glass and impinges on the air at the angle of incidence $\theta = \alpha$ undergoes total reflection with a phase-difference $\Delta = \frac{\pi}{4}$ between the two vibration components. By (108)

this occurs for any definite sort of glass at two angles, either of which may be chosen.

If we make the ray of light which was assumed at the beginning of this section to vibrate in the direction ϕ with respect to the x-axis, fall normally from below on the base of the surface at A (Fig. 5), the equation (107) will hold for the ratio of the amplitudes before and after the entrance of the ray into the parallelepiped. The first reflection at O on the oblique bounding surface then occurs at the angle of incidence α, with a phase-difference of $\frac{\pi}{4}$ and unaltered amplitudes; the second reflection occurs at P on the opposite bounding surface with a further phase-difference of $\frac{\pi}{4}$ and again unaltered amplitudes; finally, at B the ray emerges normally into the air with uniformly diminished amplitudes and unchanged phase-difference $\frac{\pi}{4} + \frac{\pi}{4} = \frac{\pi}{2}$. Hence the ray which emerges at the top at B performs elliptic vibrations, whose principal axes coincide with the co-ordinate axes x (towards the right) and y (towards the back of the figure).

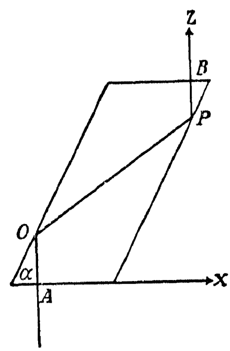

FIG. 5.

Since the light which is incident at A is assumed to be vibrating in the first and third quadrant $\left(0 < \phi \leq \frac{\pi}{4}\right)$, the light which emerges at B is lævorotatory, by (105). But if the light vibrates in the second and fourth quadrants $\left(\frac{3\pi}{4} \leq \phi < \pi\right)$ the emergent light is dextrorotatory. When $\phi = \frac{\pi}{4}$ or $\frac{3\pi}{4}$ we have $\tan \phi = \pm 1$ and $a = b$, and then the light executes left-handed or right-handed circular vibrations.

§ 26. Our discussion so far has been entirely restricted to a definite radian frequency ω—that is, to a single

partial vibration in the Fourier series of the optical vibration. Now since even the most nearly homogeneous light wave in nature includes a great number of partial vibrations, we have in all cases to sum up over all these partial vibrations and so arrive at the relevant laws for the resultant vibration.

Instead of this, however, we may also, as in (73), regard every homogeneous optical vibration as a single elliptic vibration of the corresponding frequency with the lengths and directions of the axes changing slowly and irregularly, or as consisting of two circular vibrations in opposite directions whose radii change slowly and irregularly.

But since the laws which have been deduced in the last sections, from § 22 onwards, are quite independent of the particular value of the frequency ω, they also apply unaltered when many partial vibrations of almost the same frequency are superposed. This holds, in particular, for the ratio of the amplitudes (107) of the two vibration components in a partial wave of linearly polarized light, as well as for the phase-difference Δ between the two vibration components, which is given by (108), and hence also for the direction (111) and the ratio of the lengths of the two principal axes of the resultant elliptic vibration (cf. end of § 22). All these quantities are the same for the different partial waves—that is, they are subject to no fluctuations. Optical vibrations whose direction of axes and ratio of axes are constant are said to be "elliptically polarized." This does not only express, then, that the optical vibrations are elliptic—that is obvious—but also that the vibration ellipse, even if it changes slowly and irregularly, nevertheless retains the directions of its axes, the ratio of its axes, and its sense of traverse unchanged throughout. In the same way "circularly polarized" light is represented by circular vibrations whose radius alters slowly and irregularly. Elliptically or circularly polarized light may be produced from linearly polarized light by means of Fresnel's rhomb, as above described.

The direction of the axes is determined, as we saw, by the plane of incidence at total reflection, and the ratio of the axes by the azimuth of polarization. From this we easily see that the two vibration components of light of this kind are completely coherent (§ 20).

§ 27. Taking these results as a basis we are now in a position to analyse up to a certain point any plane wave that we may encounter. Let us take any light-wave present in nature, whose origin is unknown to us, and undertake the problem of determining its form by making certain measurements, that is, to specify as far as possible the properties of the two wave-functions f and g of which it is composed.

First we resolve the light spectrally by one of the methods of interferometry above described. We are then able to deal only with homogeneous light.

We then determine, by § 11, the two principal directions of vibration and the two principal intensities for a homogeneous light-wave of this kind. Three cases can occur. The first is that one of the principal intensities vanishes entirely. This means that the light is linearly polarized. The second is the opposite extreme where the two principal intensities are equal to each other. The light is then either circularly polarized or is natural light or a mixture of both (partially circularly polarized light). To distinguish between these three possibilities we allow the light to fall normally on a Fresnel rhomb (Fig. 5). If the light is circularly polarized, the amplitudes of the two vibration components f and g are equal to each other and their phase-difference then amounts to $\frac{\pi}{2}$ or $\frac{3\pi}{2}$ according as the light is lævorotatory or dextrorotatory. After emergence from the glass the amplitudes are still equal to each other but the phase-difference now amounts to π or 2π, respectively—that is, the emergent light is linearly polarized and vibrates in the second and fourth or first and third quadrants, respectively, at the angle $\phi = \frac{3\pi}{4}$ or

$\frac{\pi}{4}$, to the x-axis. Thus here the double reflection supplements the original phase-difference of the two components to form a whole multiple of π—that is, the Fresnel rhomb acts as a "compensator."

On the other hand, if we are dealing with natural light, no compensation occurs, and the emergent light, like the original light, shows no trace of differences in the different directions of vibration.

If, lastly, the light is a mixture of circularly polarized and natural light, the natural part remains unchanged, whereas the circular part again yields linearly polarized light whose azimuth is $\frac{3\pi}{4}$ or $\frac{\pi}{4}$, respectively, and whose intensity is represented by the difference in the principal intensities of the emergent light.

In the third and most general case in which the two principal intensities differ from zero and from each other, we allow the light to fall on the rhomb in such a way that the direction of the greater principal intensity is coincident with the x-axis (Fig. 5). If the light is elliptically polarized the two vibration components have the phase-difference $\frac{\pi}{2}$ or $\frac{3\pi}{2}$, respectively, and the ratio of the amplitudes is equal to $\frac{b}{a}$. Owing to the action of the compensator, linearly polarized light is produced, which vibrates in the direction $\phi = \mp \tan^{-1}\frac{b}{a}$ and forms an angle $<\frac{\pi}{4}$ with the x-axis. But if the light is a mixture of linearly polarized and natural light, the greater principal axis also coincides with the x-axis in the case of the emergent light. If, finally, the light is partially elliptically polarized the greater principal axis in the emergent light forms an angle $<\frac{\pi}{4}$ with the x-axis, which indicates the ratio of the axes of the vibration ellipse, while the intensity of the polarized portion is obtained from the difference in the principal intensities.

In this way we obtain in every case that presents itself a certain insight into the conditions of polarization of a given ray. There is still, of course, a considerable remainder which cannot be defined and which is caused by the physical complexity of the elementary processes that occur in every light source; it receives expression in the numerous terms of the Fourier expansion.

CHAPTER III

GEOMETRICAL OPTICS

§ 28. THE laws of reflection and refraction which we have developed above owe their simple character essentially to our assumption that both the wave-points of the light (§ 5) and the boundary planes of the active body are infinite planes. In reality these surfaces are neither unlimited nor plane, and hence, strictly speaking, the above simple laws are not applicable to nature at all. Nevertheless, in practice they represent an extraordinarily close approximation to reality, when these assumptions are very nearly fulfilled, valid, that is, when both the cross-sections of the wave-fronts and the boundary planes, as well as the radii of curvature of these surfaces, are very great in comparison with the wave-length of the light under consideration. Now since in optics the wave-lengths involved are, in general, of a smaller order of magnitude than the dimensions of the bodies used, the optical laws of propagation, reflection and refraction assume a particularly simple form in contrast with acoustics where, it is true, the laws of wave-motion also hold, but where this assumption is not in general fulfilled.

Geometrical optics comprises the relationships which are obtained when the wave-fronts of the light and also the surfaces of bodies are imagined to be divided up into many small parts, and the laws of propagation, reflection and refraction of plane waves of unlimited extent which fall on unlimited plane surfaces are applied to these individual small parts. Each of these small portions of wave-front, which are regarded as plane, is graphically

represented by the corresponding ray which is normal to it, and every ray propagates itself rectilinearly with the velocity characteristic for that body until it reaches another body where it is reflected and refracted as at a plane surface in accordance with the relative position of the normal. Hence geometric optics is also called "ray optics" in contrast with "wave optics," which is more general but also more complicated.

§ 29. We take as our first example the course of a ray of light through a *prism*—that is, through a transparent body which is bounded by two plane faces which are inclined to each other at an angle ϕ. After what has been said above, we must exclude processes which occur in the immediate neighbourhood of the edge A (which in Fig. 6 we take perpendicular to the plane of the diagram), because the radius of the refracting surface becomes infinitely small there.

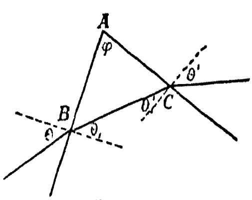

FIG. 6.

A ray which is incident at B at an angle θ in the plane of the diagram is refracted along BC and emerges into the air again at C at the angle θ' to the normal at C. The following relations hold for these two refractions :

$$\sin\theta = n\sin\theta_1 \quad . \quad . \quad . \quad . \quad . \quad (113)$$

$$\sin\theta' = n\sin\theta'_1 \quad . \quad . \quad . \quad . \quad . \quad (114)$$

Moreover, the sum of the three angles of the triangle ABC gives :

$$\left(\frac{\pi}{2} - \theta_1\right) + \phi + \left(\frac{\pi}{2} - \theta'_1\right) = \pi$$

that is :

$$\theta_1 + \theta'_1 = \phi \quad . \quad . \quad . \quad . \quad . \quad . \quad (115)$$

From these equations we can find how the angle of emergence θ' depends on the angle of incidence θ.

The total deviation δ of the ray caused by the prism—

that is, the angle between the emergent and the originally incident ray—is given by:

$$\delta = (\theta - \theta_1) + (\theta' - \theta'_1)$$

or, in view of (115), by:

$$\delta = \theta + \theta' - \phi \quad . \quad . \quad . \quad . \quad (116)$$

If we investigate the manner in which the total deflection δ depends on the original angle of incidence θ, we get, by using (113), (114) and (115):

$$\frac{d\delta}{d\theta} = 1 - \frac{\cos\theta \, . \cos\theta'_1}{\cos\theta_1 \cos\theta'} \quad . \quad . \quad . \quad (117)$$

This expression vanishes if $\theta'_1 = \theta_1$, which simultaneously makes $\theta' = \theta$—that is, when the ray traverses the prism symmetrically—so that the lengths AB and AC are equal to one another. In this case the deflection δ_0 is a maximum or a minimum, and we find out which it is by finding the value of the second differential coefficient. This is obtained by differentiating (117) with respect to θ; for the value $\theta' = \theta$, which we are considering, it has the value:

$$\left(\frac{d^2\delta}{d\theta^2}\right)_0 = \frac{2\sin(\theta + \theta_1) \, . \sin(\theta - \theta_1)}{\sin\theta \cos\theta \cos^2\theta_1} \quad . \quad (118)$$

If the index of refraction of the substance of the prism $n > 1$, then in the symmetrical position $\theta' = \theta$ the second differential coefficient is positive and hence the deflection is a minimum.

This limiting position can be determined fairly sharply experimentally by leaving the position of the incident ray unchanged, but rotating the prism to and fro on some axis parallel to its edge and observing the direction of the emergent ray. If the minimum deviation δ_0 has been found by trial, the refractive index of the substance of the prism is calculated from (116), (113) and (115) and comes out as:

$$n = \frac{\sin\dfrac{\phi + \delta_0}{2}}{\sin\dfrac{\phi}{2}} \quad . \quad . \quad . \quad . \quad (119)$$

§ 30. Before passing on to consider curved wave-fronts and curved surfaces of bodies we must make some remarks of a general character. The primary wave-front in an isotropic body is a spherical surface. For every point-like source of light sends out spherical waves of light which can be represented at a sufficient distance from the source by a beam of divergent rays. If these rays are refracted or reflected at the curved surface of another body, their directions will, in general, no longer intersect at a point—that is, they will no longer be homocentric but "astigmatic." If, in a special case, they *are* again homocentric, then the point at which they meet is called the "optical image" of the source of light, being real or virtual according as the rays converge towards or

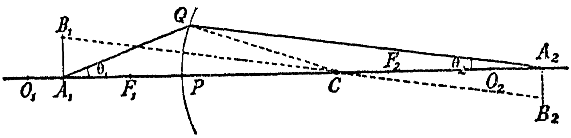

FIG. 7.

diverge from the point of intersection. Since light can always traverse its path in the reverse direction (end of § 10), the source and the image can exchange places, and are therefore also called "conjugate" points. The image of an infinitely distant point of light, or, what amounts to the same thing, the point of intersection of rays which result through refraction and reflection from an incident beam of parallel rays, is called a "focus"—that is, when it exists. If, conversely, we make the focus a source of light, we obtain a beam of parallel rays. It is on this circumstance that we base the experimental production of plane waves of light.

§ 31. We shall now investigate the refraction of light at the *spherical* bounding surface between two bodies 1 and 2 (Fig. 7). Let n_1 and n_2 be the refractive indices of the two bodies, C the centre of the curved surface, and $r = PC$, the radius of the sphere. We first consider the image of a point of light A_1 situated on the straight line

PC, the axis of the system. Of the rays which emerge from A_1 only the axial ray A_1P passes through the spherical surface without being deflected. Let another ray which emerges from A_1 at an angle θ_1 to the axis meet the spherical surface at Q. It becomes refracted there according to the law:

$$\frac{\sin A_1QC}{\sin A_2QC} = \frac{n_2}{n_1}$$

and meets the axis at the point A_2 at an angle θ_2. If we calculate the two sines from the triangles A_1QC and A_2QC we get the relation:

$$\frac{A_1C}{A_1Q} : \frac{A_2C}{A_2Q} = \frac{n_2}{n_1} \quad . \quad . \quad . \quad . \quad (120)$$

In general the position of the point A_2 will change with the value of θ_1 or with the point Q at which the ray impinges—that is, the beam which starts from A_1 becomes astigmatic after the refraction. But if θ_1 is taken sufficiently small, we can replace the lengths A_1Q and A_2Q in (120), except for errors of the second order, by the distances A_1P and A_2P. We then obtain, if we denote the distances of the points A_1 and A_2 from P by e_1 and e_2:

$$\frac{e_1 + r}{e_1} : \frac{e_2 - r}{e_2} = \frac{n_2}{n_1} \quad . \quad . \quad . \quad . \quad (121)$$

or:

$$\frac{n_1}{e_1} + \frac{n_2}{e_2} = \frac{n_2 - n_1}{r} \quad . \quad . \quad . \quad . \quad (122)$$

Since the point Q no longer occurs in the latter expressions, the beam of rays which starts out from A_1 is homocentric even after the refraction, provided the conical angle of the beam is small; and the point A_1 now has an optical image, the conjugate point A_2. By (122) the law for the formation of the image is expressed by a linear equation between the reciprocals of the distances of the conjugate points from the refracting spherical surface.

If A_1 moves off to infinity on the left, A_2 moves to the focus F_2, whose abscissa is :

$$f_2 = \frac{n_2}{n_2 - n_1} \cdot r \qquad . \qquad . \qquad (123)$$

If A_1 moves to the right, A_2 first moves off to infinity on the right—namely, until A_1 reaches the focus F_1, whose abscissa is :

$$f_1 = \frac{n_1}{n_2 - n_1} \cdot r \qquad . \quad . \quad . \quad . \quad . \qquad (124)$$

If A_1 moves beyond F_1, the refracted ray becomes divergent—that is, A_2 appears on the left as a virtual image, which follows the light A_1 as it moves further to the right, until it catches up A_1 at the point P. The point P is its own conjugate, just like the point C, since the rays which start from C are not deflected at all.

The abscissæ f_1 and f_2 of the foci exhibit the simple relationships :

$$f_2 - f_1 = r \qquad . \quad . \quad . \quad . \quad . \quad . \qquad (125)$$

$$\frac{f_1}{f_2} = \frac{n_1}{n_2} \qquad . \quad . \quad . \quad . \quad . \quad . \qquad (126)$$

If we substitute for r and n from them, the law of image formation (122) becomes simply :

$$\frac{f_1}{e_1} + \frac{f_2}{e_2} = 1 \qquad . \quad . \quad . \quad . \quad . \quad . \qquad (127)$$

All these relations, as we shall see, are capable of being considerably generalized. We may add here that the formation of an image by reflection at the spherical surface may be included in the above by writing $n_1 = - n_2$.

If the rays which start out from A_1 intersect at the image point A_2 after refraction, the angular aperture or the " divergence " of a beam at A_1 is different from that of the conjugate beam at A_2. For from Fig. 7 we get for the angle of inclination of the two conjugate rays A_1Q and A_2Q to the central axis :

$$\sin \theta_1 : \sin \theta_2 = A_2Q : A_1Q$$

and to a corresponding degree of approximation:

$$\sin \theta_1 : \sin \theta_2 = A_2P : A_1P = e_2 : e_1 \quad . \quad . \quad (128)$$

§ 32. We shall now investigate the image of a point of light which does not lie on the central axis. Then B_1 has an optical image, just as much as A_1, namely the conjugate-point B_2 which lies on the straight line B_1C. The circle which results from rotating B_1 about the central axis A_1A_2 forms an image which is the circle described by the rotation of B_2, and the circular area of radius A_1B_1 forms its image on a certain surface of rotation which is bounded by the circle B_2 and which has its centre at A_2. If we restrict ourselves to distances which are not far from the central axis, we may regard this surface of rotation as a plane of circular contour (radius A_2B_2) which is perpendicular to the central axis. The surface of light then forms a "similar" image on the image plane, and the following equation holds for the ratio of the length of the light-path $A_1B_1 = l_1$ to the length of the image-line $A_2B_2 = l_2$:

$$\begin{aligned} l_1 : l_2 &= A_1C : A_2C \\ &= (e_1 + r) : (e_2 - r) \end{aligned}$$

and by (121):

$$l_1 : l_2 = \frac{n_2 e_1}{n_1 e_2}.$$

Together with (128) this gives the following simple relationship between the ratio of the linear dimensions and the ratio of the ray directions in two conjugate planes:

$$n_1 l_1 \sin \theta_1 = n_2 l_2 \sin \theta_2 \quad . \quad . \quad . \quad . \quad (129)$$

If the point A_1 coincides with the focus F_1, $l_2 = \infty$ and $\theta_2 = 0$. The two planes which pass through the two foci and are perpendicular to the axis are called the "focal planes."

§ 33. All the above laws can easily be generalized for the case of any arbitrary number of refracting spherical surfaces. We only need to choose the measures for the

abscissæ of two conjugate points A_1 and A_2 appropriately. Hitherto we have used for the abscissæ their distances e_1 and e_2 from the refracting spherical surface, and we have reckoned the distances e_1 as positive towards the left and the distances e_2 as positive towards the right. The origin P is conjugate to itself. We shall now choose a special origin for each of the two abscissæ, but in such a way that the two origins O_1 and O_2 are again conjugate to each other, but otherwise arbitrary. Further, we shall assume the directions of the abscissæ x_1 and x_2 to be positive towards the right in both cases. If a_1 and a_2 then denote the new origins O_1 and O_2 of the spherical surface, where by (127):

$$\frac{f_1}{a_1} + \frac{f_2}{a_2} = 1 \quad . \quad . \quad . \quad . \quad . \quad (130)$$

then the relations between the new abscissæ x_1 and x_2 and the old abscissæ e_1 and e_2 are (Fig. 7):

$$O_1A_1 = O_1P - A_1P = a_1 - e_1 = x_1$$
$$O_2A_2 = PA_2 - PO_2 = e_2 - a_2 = x_2$$

and we get for the law of image formation from (127), if we replace the e's in it by the x's, and take (130) into account:

$$\frac{a_1 - f_1}{x_1} + \frac{f_2 - a_2}{x_2} = 1 \quad . \quad . \quad . \quad (131)$$

This is again a linear equation between the reciprocals of the abscissæ of the two conjugate points.

§ 34. We now pass on to investigate a system consisting of an *arbitrary number of line-centred spherical surfaces*, that is, spherical surfaces whose centres all lie on a straight line, the central axis of the system. If we again restrict our attention to such rays and such points of light as lie near the central axis, that is, if we consider only small values of θ and l, the laws of image formation may be derived directly from those obtained above. Let us first take a point of light A situated on the axis and a

narrow beam of light which starts out from it in the first body. Corresponding to it there is a definite conjugate point on the axis and a definite conjugate beam of rays (real or virtual) in the second body; and corresponding to these again there are definite conjugate quantities in the third body. Proceeding in this way we finally arrive at a definite conjugate point A' and a definite conjugate beam of rays in the last body.

Now since the reciprocals of the abscissæ of two successive conjugate points, when referred to any two points as origins, depend linearly on each other, there must also be a linear relationship between the abscissæ x and x' of the conjugate points A and A' in the first and last bodies, and the law of image-formation has a form similar to (131), namely:

$$\frac{f}{x}+\frac{f'}{x'}=1 \quad . \quad . \quad . \quad . \quad . \quad (132)$$

For $x=0$, x' becomes equal to 0. A and A' then coincide with the origins O and O'. The constants f and f' are clearly the abscissæ of the foci in the first and the last bodies. If the foci lie on different sides of the origins, f and f' have opposite signs.

By (129) the following relation holds for the lengths l and l' of two conjugate lines that pass through A and A' and are normal to the central axis:

$$nl\sin\theta = n'l'\sin\theta' \quad . \quad . \quad . \quad . \quad (133)$$

where n and n' denote the refractive indices, θ and θ' the angles of inclination which any two conjugate rays that pass through A and A' make with the central axis.

In the sequel we shall reckon both the lengths l and l' and also the angles θ and θ' as positive in the same direction: this causes no change of sign in (133).

§ 34*a*. Let us now see whether there are two conjugate points for which $l=l'$, that is, for which the representation of the two surfaces that pass through them normally to the central axis is fully congruent. We shall call these

two points H and H' and shall try to find their abscissæ x and x'. From (133) we then get:

$$n \sin \theta = n' \sin \theta'$$

or, if we draw any two conjugate rays through H and H' which make the (small) angles θ and θ' with the central axis and meet the planes that pass through the origins O and O' at the distances l_0 and l'_0 from the origins (Fig. 8), we have:

$$n \frac{l_0}{x} = n' \frac{l'_0}{x'}$$

From this and (132) we get:

$$\left.\begin{aligned} x &= f + \frac{nl_0}{n'l'_0} f' \\ x' &= \frac{n'l'_0}{nl_0} f + f' \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (134)$$

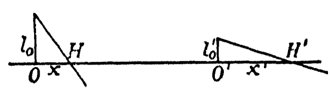

FIG. 8.

The points H and H' which are uniquely defined by these equations are called the "principal points" and the planes which pass through them and are congruent representatives of each other are called the "principal planes" of the system.

The results become particularly simple if we make the origins O and O' coincident with the principal points H and H'. For then x and x' become equal to zero in (134), and $l_0 = l'_0$, consequently:

$$f : f' = -n : n' \quad . \quad . \quad . \quad . \quad (135)$$

that is, the foci lie on opposite sides within or without the principal points, and their distances from the principal points, the so-called "focal distances," are in the ratio of the refractive indices. We then get for the law of images (132):

$$\frac{n}{x} - \frac{n'}{x'} = \frac{n}{f} \quad . \quad . \quad . \quad . \quad . \quad (136)$$

If the first body is of the same nature as the last, for example, air, then $n' = n$ and the focal distances become equal, so that the law of images becomes simplified to :

$$\frac{1}{x} - \frac{1}{x'} = \frac{1}{f} \quad . \quad . \quad . \quad . \quad . \quad (136a)$$

§ 35. Analogous to the two principal points H and H' whose planes form congruent images there are two conjugate points K and K', the nodal points, whose beams form congruent images, so that for each pair of conjugate rays $\theta = \theta'$. Their position is obtained directly by observing that two such conjugate rays which pass through K and K' and are parallel to each other intersect the principal planes at the same distance $l_0 = l'_0$ from the central axis; that is, the abscissæ x and x' of the two

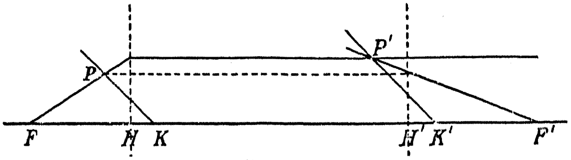

FIG. 9.

nodal points are also equal to each other. Hence it follows from (132) that :

$$x = x' = f + f' = f\left(1 - \frac{n'}{n}\right) \quad . \quad . \quad . \quad . \quad (137)$$

that is, the nodal points are at the same distance from the principal points, namely a distance equal to the difference of the focal lengths.

The relative position of the principal points, foci and nodal points are depicted graphically in Fig. 9, in which the abscissæ of the nodes are positive and the foci lie outside the principal points, so that $f + f' > 0$, $f < 0$, $f' > 0$. Then :

$$HK = H'K' = H'F' - HF$$

Consequently :

$$H'F' = HK + HF = FK$$

that is, the point-pairs FF', HK' and KH', which, taken together, are also called the "cardinal points," have a common centre, the so-called optical centre of the system. If there is only a single refracting surface (Fig. 7), this surface forms a congruent image on itself; that is, the two principal planes coincide in it. Consequently the nodal points also coincide, namely with the other self-conjugate point, the centre C of the sphere, as is immediately clear, since every ray which passes through C is self-conjugate. This causes the equations (135) and (137) for the ratio and the difference of the focal distances to transform directly into the special equations (126) and (125) which were found earlier in § 31.

If we have a system of glass lenses in air, the focal distances are equal to each other and the nodal points coincide with the principal points. If only a single thin lens is present, the principal points coincide at a point within the lens.

In general, the position of the cardinal points leads directly to a simple geometrical method of determining the point P' in the last body which is conjugate to any point P in the first body (Fig. 9). We need only draw the conjugate rays to any two rays that pass through P. The simplest ray to use is PK which passes through the nodal point K. The parallel ray $P'K'$ is conjugate to it. Besides this, we can also take the ray through P parallel to the axis. This ray has as its conjugate that ray in the last body which passes through the focus F' in it and intersects the principal plane in it at the same distance from the central axis as that at which the original ray intersected the principal plane in the first body. This serves to define P'. In addition, although this is unnecessary, we may also consider the ray PF. This ray has as its conjugate a ray, parallel to the axis, which intersects the last principal plane at the same distance from the axis as the first ray does the principal plane.

§ 36. A complete description of the laws of optical image formation must deal not only with the course of

the rays but also with the intensity of the rays. We must above all bear in mind here that a finite quantity of radiant energy never starts out from a point but always from a surface, and never in a definite direction but always within a cone of directions. We can depict this circumstance graphically by representing the energy of a beam of rays by means of the number of rays contained in the beam. Every beam then contains a quadruply infinite number of rays, since the surface from which it starts out defines a doubly infinite number of points, and each point a cone comprising a doubly infinite number of directions. But there is a certain difference which must not be overlooked. The rays which start out in various directions from a luminous point are coherent with one another, corresponding to the circumstance that the point is the centre of a wave-surface, every point of which is in the same phase. But the rays which come from two different points of the initial surface are in general non-coherent, particularly when the initial surface acts as the source of light. Hence in general no interference occurs between such rays; their intensities simply become added. But it may happen—and this is sometimes overlooked—that two different points of a surface emit coherent light; and again, conversely, non-coherent rays may start out from a single point of the surface. For example, when the surface (aperture or slit) is illuminated by light from another source; in this case certain interference phenomena can occur for which geometrical optics taken alone cannot account. We shall therefore always assume that the initial plane is self-luminous.

If the source of light is placed in a focal plane, a beam of parallel rays results, but, in conformity with the above remarks, not in the sense that all rays of the beam are parallel, but rather that, corresponding to every individual point of the luminous surface there is a particular direction of the rays of the beam, whereas the size of the luminous surface determines the angle of aperture of the cone of rays which emerge in various directions. Con-

versely, rays which come from a practically infinitely distant source, such as the sun, will not all become focused at one point; rather, a small picture of the sun will be produced in the focal plane, since all the parallel rays which come from a definite point of the sun meet at a definite point in the focal plane.

The intensity J of the radiation emitted normally by a small surface element f perpendicular to the central axis, the rays lying within a narrow cone of angular aperture Ω is, after what has been said above :

$$J = K \,.\, f \,.\, \Omega \quad . \quad . \quad . \quad . \quad . \qquad (138)$$

where the finite quantity K is called the "specific intensity of radiation" or the "specific luminosity" of the beam.

In the same way the following relation holds for the conjugate beam which lies within the conjugate cone of rays Ω' and is incident on the conjugate surface f' :

$$J' = K' \,.\, f' \,.\, \Omega' \quad . \quad . \quad . \quad . \quad . \qquad (139)$$

By the laws of image formation there is a general relationship between the quantities f and Ω in the first and the last body, which is fixed by (133). Since, on account of the similarity of the image with the object, the form of the surface f is of no account, we may choose it to be a circle of radius l. We then have

$$f = l^2\pi, \; f' = l'^2\pi.$$

If, further, we imagine Ω to be a circular cone with a small angle of aperture θ, then.

$$\Omega = 4\pi \sin^2 \frac{\theta}{2} = \theta^2\pi.$$

In the same way :

$$\Omega' = \theta'^2\pi.$$

These values, combined with (138) and (139), give the relation :

$$n^2 f\Omega = n'^2 f'\Omega' \quad . \quad . \quad . \quad . \quad . \qquad (140)$$

and, in consequence of (138) and (139):

$$\frac{n^2 J}{K} = \frac{n'^2 J'}{K'} \quad . \quad . \quad . \quad . \quad . \quad (141)$$

If the total intensity of radiation J is the same in the two conjugate beams—which will, however, never actually be the case, if only because the radiation suffers a loss at every refraction—we have:

$$\frac{K}{n^2} = \frac{K}{n'^2} \text{ or } Kq^2 = K'q'^2 \quad . \quad . \quad . \quad (142)$$

That is, the intensities are in the inverse ratio of the squares of the velocities of transmission of the rays. For $n = n'$ we get $K = K'$. Hence the intensity of a beam of rays can never be increased by optical image formation alone. It is true that the radiation is concentrated on a smaller surface at the focus but then it diverges in various directions.

§ 37. The simple laws above developed for optical image formation by means of a system of line-centred spherical surfaces mostly hold only as a first approximation to the real conditions. For in practical optics we deal both with surfaces of finite size as well as with finite cones of rays. In addition to the corrections necessitated by these circumstances there is another due to the fact that the refractive index depends on the colour of the light (§ 9). If all these influences are to be taken into account, problems of a difficult kind arise; their solution has played the predominant part in the highly developed technique of construction of optical instruments. But the laws of geometrical optics, such as we have described them in § 28, are not affected by these practical problems. They may all be condensed into a single simple graphical law, which is one of the first examples of the equivalence of an integral law with a differential law (cf. III, § 42), namely, *Fermat's Principle of Least Time.*

This principle controls the course of a ray of light through any arbitrary number of different bodies with

any boundary surfaces by asserting that the time which light requires to pass from any definite point on a ray to another definite point on the same ray is less along the actual path of the ray than along any neighbouring path.

A simple proof of this principle is obtained if we consider the optical wave-surface that starts out from any luminous point P as centre. According to § 5 there is a definite wave-front corresponding to every moment of time t. It signifies the boundary to which light which is emitted by P at the time $t = 0$ has advanced in all directions during the interval of time 0 to t, independently of the presence of any bodies whatsoever in the vicinity of P. Thus the wave-front comprises the end-points of all the rays which start out from P in all directions, each propagating itself along its own particular path. Now if Q is any point on the wave-surface, PQ denotes a definite ray, and the light takes the time t to propagate itself from P along the path of the ray to Q. Now if the light from P were to take a smaller time than t to pass from P along any neighbouring path to Q, then the propagation of light from P in all directions would proceed beyond Q in the interval 0 to t, and this contradicts the definition of wave-surface. Consequently the time t is the minimum of all the times taken by the light to arrive from P to Q along any path.

The whole of geometrical optics can be developed from Fermat's principle, and even for non-homogeneous and anisotropic bodies. The result that a ray travels in a straight line in a homogeneous isotropic medium follows directly, according to Fermat's principle, from the circumstance that the straight line is the shortest distance between the two points P and Q.

Let us consider, as a second example, the passage of light through the plane boundary surface of two homogeneous isotropic bodies. If q and q_1 are the velocities of propagation, A and B the projections of P and Q on the boundary, C the point through which the ray PQ

passes through the boundary (Fig. 10), and θ and θ_1 the angles of incidence and refraction, then the time which light takes to pass from P to Q is :

$$t = \frac{PC}{q} + \frac{CQ}{q_1} = \frac{r}{q} + \frac{r_1}{q_1} \quad . \quad . \quad . \quad (143)$$

where :

$$r^2 = AP^2 + AC^2, \text{ and } r_1^2 = BQ^2 + BC^2 \quad . \quad (144)$$

If the point C through which the ray crosses the boundary is displaced, the distance AP, BQ and AB remains constant, whereas $AC = a$ and $BC = b$ change, so that :

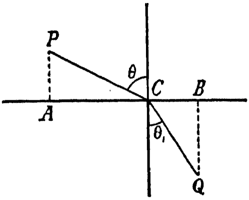

FIG. 10.

$$\delta a + \delta b = 0 \quad . \quad . \quad . \quad . \quad . \quad (145)$$

Consequently, by (143) and (144) :

$$\delta t = \frac{a\delta a}{rq} + \frac{b\delta b}{r_1 q_1}$$

If we set $\delta t = 0$ and take (145) into account, it follows that :

$$q : q_1 = \frac{a}{r} : \frac{b}{r_1} = \sin\theta : \sin\theta_1$$

which is identical with (20).

Fermat's principle may be used in a similar way to find the direction of propagation of light through any arbitrary bodies with any arbitrary surfaces.

If the end-point Q of a ray which starts out from a luminous point P coincides with a point which is optically conjugate to P the ray-path PQ is not definite ; rather, there are an infinite number of rays from P to Q. Consequently the time which the light takes to pass from P to Q is the same along all these paths, and the optical wave-surface around P has a singular point in Q.

CHAPTER IV

DIFFRACTION

§ 38. If the wave-fronts or the surfaces of the bodies may no longer be regarded as of infinite extent compared with the wave-lengths, the laws of geometrical optics fail, the conception of a ray and its propagation lose their meaning, and phenomena occur which are denoted by the term "diffraction." It is clear from the discussion in § 28 that the part played by diffraction becomes so much the greater the greater the wave-length of the light used. An exact theory of diffraction is possible only on the basis of, firstly, the differential equations for electromagnetic waves in the medium under consideration, which we shall always take to be a vacuum space in which light has the velocity $q = c$, and, secondly, the boundary conditions which hold at the surfaces of the diffracting bodies.

The differential equations for the interior can always be reduced to a single differential equation (III, § 87), which must be satisfied by every component ϕ of every vector. It is the wave-equation:

$$\ddot{\phi} = c^2 \, . \, \Delta\phi \quad . \quad . \quad . \quad . \quad . \quad . \qquad (146)$$

The boundary conditions differ according to the nature of the diffracting body. They come out most simply for so-called absolute conductors (III, § 92), in which the electric intensity of field vanishes and which act as perfect mirrors. The opposite extreme is given by the so-called black bodies which absorb all the incident light. Diffraction problems are among the most difficult encountered in optics. The exact solution of a problem of this kind was first given by Arnold Sommerfeld in 1895.

Since the wave-function ϕ can always be represented as a Fourier series, the general solution of (146) can be reduced to a sum of particular solutions, each of which corresponds to a definite frequency ω. We may write as a particular solution of this kind :

$$\phi = \psi \cdot e^{i\omega t} \quad . \quad . \quad . \quad . \quad . \qquad (146a)$$

where ψ denotes a certain complex function of the space-co-ordinates, whose absolute value (modulus) characterizes the amplitude and whose argument is the phase-constant of the periodic motion (cf. § 18 above).

The wave-equation (146) then leads to the following differential equation of the function ψ in space-co-ordinates :

$$\Delta\psi + \frac{\omega^2}{c^2}\psi = 0$$

or :

$$\Delta\psi + \frac{4\pi^2}{\lambda^2}\psi = 0 \quad . \quad . \quad . \quad . \qquad (146b)$$

which is to be integrated with due regard to the boundary conditions.

Now as the rigorous solutions of the diffraction problem formulated in this way are of a very intricate character and demand a relatively heavy mathematical equipment, it is the more important that approximate solutions, which are fully sufficient for the short optical waves, should be obtained by introducing *Huygen's Principle.* This principle is founded on the idea that every point of space on which a light-wave impinges itself becomes the centre of a new light-wave which spreads out from it in concentric spherical surfaces. G. Kirchhoff was the first to succeed in formulating Huygen's principle exactly ; it forms an extension of Green's theorem in the theory of potential, which we proceed to deduce here.

§ 39. Let U and V be two uniform functions of the space-co-ordinates x, y, z. These functions, as well as their differential coefficients, are to be continuous. Then for a space whose volume-element is $d\tau$, surface-

element $d\sigma$ and inwardly directed normal ν we have, by II, § 18, equation (80):

$$\int\left(\frac{\partial U}{\partial x}\frac{\partial V}{\partial x}+\frac{\partial U}{\partial y}\frac{\partial V}{\partial y}+\frac{\partial U}{\partial z}\frac{\partial V}{\partial z}\right)d\tau=-\int V\frac{\partial U}{\partial \nu}d\sigma-\int V\Delta U d\tau$$
$$=-\int U\frac{\partial V}{\partial \nu}d\sigma-\int U\Delta V d\tau.$$

Consequently:

$$\int\left(U\frac{\partial V}{\partial \nu}-V\frac{\partial U}{\partial \nu}\right)d\sigma=\int(V\Delta U-U\Delta V)d\tau \quad . \quad (147)$$

If we set $U=\frac{1}{r}$, where r denotes the distance (taken as positive) of the point x, y, z from any point A which lies outside the region of integration, then U fulfils the conditions of uniformity and continuity; further, we have Laplace's equation I (129) $\Delta U=0$. Consequently (147) becomes:

$$\int\left(\frac{1}{r}\frac{\partial V}{\partial \nu}-V\cdot\frac{\partial\frac{1}{r}}{\partial \nu}\right)d\sigma=-\int\frac{\Delta V}{r}d\tau \quad . \quad . \quad (148)$$

If the point A lies inside the assumed space, this equation may also be applied, provided that we exclude a very small sphere with its centre at A from the integration. This does not appreciably alter the right-hand side (cf. I, § 33), but the left-hand side acquires an additional term in the form of an integral over the surface-elements $d\sigma_1$ of the sphere whose normal, directed towards the interior of the integration space, coincides with r, thus:

$$\int\left(\frac{1}{r}\frac{\partial V}{\partial r}+V\cdot\frac{1}{r^2}\right)d\sigma_1.$$

If the radius of the sphere is taken sufficiently small, this integral, since $\frac{\partial V}{\partial r}$ and V are finite, reduces to:

$$\int\frac{V}{r^2}d\sigma_1=V_0\int\frac{d\sigma_1}{r^2}=4\pi V_0$$

where V_0 denotes the value of V for $r = 0$, that is, at the point A, and when this is added to the left-hand side of the equation (148) we arrive at Green's theorem :

$$4\pi V_0 = -\int \frac{\Delta V}{r} d\tau + \int \left(V \frac{\partial \frac{1}{r}}{\partial \nu} - \frac{1}{r}\frac{\partial V}{\partial \nu} \right) d\sigma \,. \qquad (149)$$

which enables us to obtain the expression of a function V at every point A of the assumed space, if the value of ΔV is known at every point inside and the values of V and $\frac{\partial V}{\partial \nu}$ are known at every point on its surface.

If the function V satisfies Laplace's equation $\Delta V = 0$ the space integral drops out, and we require only the data for the surface. It is particularly in this form that Green's theorem is often useful, but its importance is essentially restricted in that the values of V and $\frac{\partial V}{\partial \nu}$ may not be chosen independently of each other. For we know that the expression for V is everywhere completely determined by the values of V at the surface (III, § 19) and simultaneously by the values of $\frac{\partial V}{\partial \nu}$ at the surface (II, § 71). Hence if either only V or only $\frac{\partial V}{\partial \nu}$ is given at the surface, Green's law can be successfully applied to determining V in the interior only when we have in some special way acquired a knowledge of the missing surface values.

§ 40. Huygen's principle follows directly from Green's law if we substitute for V a certain wave-function which is chosen in such a way that the space-integral in (149) can be reduced to a surface-integral. This is not, of course, made possible by simply substituting for V a solution ϕ of the wave-equation (146). We shall see, on the other hand, that it is always possible to reduce the space-integral to a surface-integral if, in any function ϕ of x, y, z, t which satisfies the wave-equation (146), we

substitute $t - \frac{r}{c}$ for t—where r, as before, denotes the distance of the point x, y, z from the point A, which is situated in the interior of the integration space—and then identify the expression so obtained with V in (149).

We must therefore be careful to note that V now depends in two ways on the co-ordinates x, y, z: in the first place, explicitly, since ϕ is a space-function, and, secondly, implicitly, since the argument:

$$t - \frac{r}{c} = \tau \quad . \quad . \quad . \quad . \quad . \quad . \qquad (150)$$

contains the distance r. To call attention to this we set:

$$V = \phi_\tau(x, y, z) \quad . \quad . \quad . \quad . \quad . \qquad (151)$$

in which we write the space co-ordinates in brackets on the line but attach the time co-ordinate as a suffix. There is no reason to fear confusion between the time-argument τ and the space-element $d\tau$, particularly as the latter will soon vanish from the formulæ. Hence we get from (149):

$$4\pi\phi_t(0) = -\int \frac{\Delta\phi_\tau}{r}\, d\tau + \int \left(\phi_\tau \frac{\partial \frac{1}{r}}{\partial \nu} - \frac{1}{r}\frac{\partial \phi_\tau}{\partial \nu}\right) d\sigma \ . \qquad (152)$$

This is the value of the wave-function ϕ at the time t at any point A of the integration space enclosed by the surface σ which has the inwardly directed normal ν. To reduce the space-integral to a surface-integral we first calculate the value of $\Delta\phi_\tau$, which may not, of course, be set equal to $(\Delta\phi)_\tau$. For this we have:

$$\frac{\partial \phi_\tau}{\partial x} = \left(\frac{\partial \phi}{\partial x}\right)_\tau + \left(\frac{\partial \phi}{\partial t}\right)_\tau \cdot \frac{\partial \tau}{\partial x} = \left(\frac{\partial \phi}{\partial x}\right)_\tau - \frac{1}{c}\left(\frac{\partial \phi}{\partial t}\right)_\tau \frac{\partial r}{\partial x}$$

and, further:

$$\frac{\partial^2 \phi_\tau}{\partial x^2} = \left(\frac{\partial^2 \phi}{\partial x^2}\right)_\tau - \frac{2}{c}\left(\frac{\partial^2 \phi}{\partial x \partial t}\right)_\tau \frac{\partial r}{\partial x} + \frac{1}{c^2}\left(\frac{\partial^2 \phi}{\partial t^2}\right)_\tau \left(\frac{\partial r}{\partial x}\right)^2 - \frac{1}{c}\left(\frac{\partial \phi}{\partial t}\right)_\tau \frac{\partial^2 r}{\partial x^2}$$

Corresponding expressions hold for $\frac{\partial^2 \phi_\tau}{\partial y^2}$ and $\frac{\partial^2 \phi_\tau}{\partial z^2}$.

By addition we get :

$$\Delta\phi_\tau = (\Delta\phi)_\tau - \frac{2}{c}\left(\frac{\partial^2\phi}{\partial x\partial t}\frac{\partial r}{\partial x} + \frac{\partial^2\phi}{\partial y\partial t}\frac{\partial r}{\partial y} + \frac{\partial^2\phi}{\partial z\partial t}\frac{\partial r}{\partial z}\right)_\tau$$
$$+ \frac{1}{c^2}\left(\frac{\partial^2\phi}{\partial t^2}\right)_\tau - \frac{2}{cr}\left(\frac{\partial\phi}{\partial t}\right)_\tau$$

If we now substitute :

$$(\Delta\phi)_\tau = \frac{1}{c^2}\left(\frac{\partial^2\phi}{\partial t^2}\right)_\tau$$

from the wave-equation (146), the expression for $\Delta\phi_\tau$ may be written as a differential coefficient with respect to the time :

$$\Delta\phi_\tau = -\frac{2}{c}\frac{\partial}{\partial t}\left(\frac{\partial\phi}{\partial x}\frac{\partial r}{\partial x} + \frac{\partial\phi}{\partial y}\frac{\partial r}{\partial y} + \frac{\partial\phi}{\partial z}\frac{\partial r}{\partial z} - \frac{1}{c}\frac{\partial\phi}{\partial t} + \frac{\phi}{r}\right)_\tau$$

or :

$$\Delta\phi_\tau = -\frac{2}{c}\frac{\partial}{\partial t}\left(\frac{\partial\phi_\tau}{\partial r} + \frac{\phi_\tau}{r}\right)$$

Using this expression we get for the space-integral in (152) :

$$\int\frac{\Delta\phi_\tau}{r}\,d\tau = -\frac{2}{c}\frac{\partial}{\partial t}\int\frac{d\tau}{r}\left(\frac{\partial\phi_\tau}{\partial r} + \frac{\phi_\tau}{r}\right)$$

and if we express the space-element $d\tau$ in terms of polar co-ordinates with the origin A and the angle of aperture $d\omega$ of an elementary cone whose vertex is at A :

$$d\tau = r^2 dr d\omega$$

$$\int\frac{\Delta\phi_\tau}{r}\,d\tau = -\frac{2}{c}\frac{\partial}{\partial t}\int dr d\omega\,.\left(r\frac{\partial\phi_\tau}{\partial r} + \phi_\tau\right)$$
$$= -\frac{2}{c}\frac{\partial}{\partial t}\int d\omega\,.\,[r\phi_\tau]_0^r \quad . \quad . \quad . \quad . \quad (153)$$

where we have now to insert in the definite integral containing the square bracket the distance r of the surface-element $d\sigma$ from A as the upper limit and $r = 0$ as the lower limit. Since the latter value vanishes, we are left with only the term referring to the surface, and this is related to the size of the surface-element $d\sigma$, cut out

from the integration space by the elementary cone $d\sigma$, as follows :

$$r^2 d\omega = -d\sigma \,.\, \cos(r,\nu) = -d\sigma \cdot \frac{\partial r}{\partial \nu} \quad . \quad (154)$$

since the directions of increasing r and increasing ν form an obtuse angle with each other. Hence it follows that :

$$\int \frac{\Delta\phi_\tau}{r}\,d\tau + \frac{2}{c}\frac{\partial}{\partial t}\int d\sigma \cdot \frac{\phi_\tau}{r}\frac{\partial r}{\partial \nu} = \frac{2}{c}\int \frac{d\sigma}{r}\left(\frac{\partial\phi}{\partial t}\right)_\tau \cdot \frac{\partial r}{\partial \nu} \quad (155)$$

If the elementary cone $d\omega$ in question cuts the surface of the integration space more than once—which can happen only an odd number of times, the integral (·153) in r resolves into several separate parts, and we can then easily convince ourselves that the relation (155) remains valid so long as we perform the integration with respect to $d\sigma$ over the whole of the surface which encloses the integration space.

If, finally, we remember that in (152) :

$$\frac{\partial\phi_\tau}{\partial\nu} = \left(\frac{\partial\phi}{\partial\nu}\right)_\tau - \frac{1}{c}\left(\frac{\partial\phi}{\partial t}\right)_\tau \frac{\partial r}{\partial \nu} \quad . \quad . \quad . \quad (156)$$

Green's theorem (152) passes over into Huygen's principle:

$$4\pi\phi_t(0) = \int \Omega_\nu d\sigma \quad . \quad . \quad . \quad . \quad (157)$$

where, in accordance with (155) and (156), we have used the abbreviation :

$$\Omega_\nu = \left(\phi\frac{\partial \frac{1}{r}}{\partial \nu} - \frac{1}{r}\frac{\partial\phi}{\partial\nu} - \frac{1}{cr}\frac{\partial\phi}{\partial t}\frac{\partial r}{\partial\nu}\right)_t \quad . \quad . \quad (158)$$

Instead of this we sometimes find written :

$$\Omega_\nu = \frac{\partial}{\partial\nu}\left(\frac{\phi_\tau}{r}\right) - \frac{1}{r}\left(\frac{\partial\phi}{\partial\nu}\right)_\tau \quad . \quad . \quad . \quad (159)$$

which is, however, wrong in this form. For this does not express the fact that in the first term the differentiation of $\phi_\tau\,(x, y, z)$ with respect to ν must be performed only with respect to the functional dependence on the argument τ and not on that of the co-ordinates x, y, z.

The importance of Huygen's principle is due to the fact that, according to it, the wave-function ϕ at any point A of space at any time t is completely determined by the values of ϕ, $\frac{\partial \phi}{\partial \nu}$ and $\frac{\partial \phi}{\partial t}$ at the surface of any space which contains A and in which the wave-equation (146) holds everywhere, these values being for certain moments of time τ which are different for every surface-element $d\sigma$ and which can easily be pictured owing to the circumstance that they precede the moment t by the time which the light takes to travel from $d\sigma$ to A.

Accordingly, every surface-element $d\sigma$ continually sends out, in consonance with the phenomena which occur in it, certain spherical waves with the velocity of light in all directions, and the value of the wave-function at A is made up of the sum of the contributions which arrive there at the time t from all directions. We must not forget, however, that this theorem is valid only if the surface σ is completely closed. There is no definite physical meaning in speaking of the effect of individual surface-elements $d\sigma$. For the integral (157) can be resolved into its differentials in an infinite number of ways.

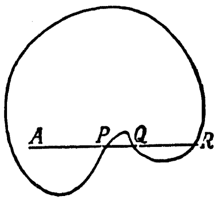

FIG. 11.

The significance of the restriction just stated becomes particularly clear in the case where, among the straight lines that start from A, there are some which intersect the surface σ in more than one point, say in the three points P, Q, R (Fig. 11). The length PQ does not then belong to the integration space and so the wave-equation (146) need not hold between P and Q; indeed, any bodies opaque to light may be situated there. Nevertheless Huygen's principle holds in the form (157); and we have to perform the surface integration over the whole of the surface σ and may not omit the surface-elements at Q and R on the plea that the light from them cannot by travelling in a straight line reach A.

§ 41. If we divide the whole integration space into two parts 1 and 2, so that the part 1 contains the point A, we obtain two spaces with the surfaces $d\sigma_1$, $d\sigma_2$ and $d\sigma_0$; we shall assume the latter two to denote the surface-elements of the surface of separation. We can then apply Huygen's principle in two different forms; firstly to the original total surface, with the surface-elements $d\sigma_1$ and $d\sigma_2$, so that:

$$4\pi\phi_t(0) = \int \Omega_{\nu_1} d\sigma_1 + \int \Omega_{\nu_2} d\sigma_2$$

and secondly to the surface of the space 1, with the surface-elements $d\sigma_1$ and $d\sigma_0$, so that:

$$4\pi\phi_t(0) = \int \Omega_{\nu_1} d\sigma_1 + \int \Omega_{\nu_{01}} d\sigma_0$$

where ν_{01} denotes the normal of $d\sigma_0$, directed inwardly towards 1.

By subtracting the last equation from the preceding one we get:

$$0 = \int \Omega_{\nu_2} d\sigma_2 - \int \Omega_{\nu_{01}} d\sigma_0$$

or, since ν_{01} and ν_{02} are opposite:

$$0 = \int \Omega_{\nu_2} d\sigma_2 + \int \Omega_{\nu} d\sigma_0 \quad . \quad . \quad . \quad (160)$$

Now, the elements $d\sigma_2$ and $d\sigma_0$ form the closed surface of space 2. Hence we have the general law that Huygen's integral, when applied to a point A outside the integration space, has the value zero—which corresponds exactly with Green's theorem in the form (148).

§ 42. We shall now apply Huygen's principle to a space which is bounded on the one hand by a surface which comprises all the bodies that are involved in the process in question, namely, sources of light, screens and so forth, on the other hand by an immensely great spherical surface of radius R and with its centre in finite space. The equation (157) is then also applicable in this case provided only that the equation (146) holds in the whole integration space. But Huygen's integral then resolves into two separate partial integrals: one which

stretches over the surface in finite space, having as its element $d\sigma$ and as its normal ν, directed towards the integration space, that is, towards the point A; the other stretching over the spherical surface at a distance R and having as its element dS. We cannot make this second integral vanish by simply choosing R very great. For even if the integrand Ω decreases to an indefinite extent as R increases, the magnitude of the surface S increases beyond all limits. Nevertheless we can always arrange that this integral can be neglected by making an appropriate assumption about the form of the wave-function ϕ—an assumption which must be always admissible and in no way restrict the actual course of the process. Namely, we need only assume that for all times $t < T$, where T denotes an immensely long period of time, the wave-function ϕ with all its derivatives everywhere vanishes; in other words, that all the sources of light started into activity at a far distant moment of time. For then the quantity ϕ_τ and its derivatives vanish for all surface-elements dS, so long as:

$$\tau = t - \frac{R}{c} < -T$$

that is, so long as we choose:

$$R > (T + t) \, . \, c.$$

It is true that the variable time t is included in this condition. But that introduces no difficulty. For we can choose R from the outset so that the inequality is also satisfied for even the greatest possible time that can occur in the measurements. In physical language, R must be chosen so great that an effect which starts out from the distant spherical surface cannot make itself felt at A at the time t.

After these remarks we may also regard the equation (157) as valid for the case where the point A lies outside the surface σ and where the bodies producing the effects all lie inside the surface.

Summarising, we may say that Huygen's principle holds in the form (157) for every case where the surface σ completely shuts off the point A from all the active bodies, so that there is no path from any of the bodies to A which does not somewhere pierce the surface σ.

§ 43. We shall formulate the general problem of diffraction which concerns us here as follows. A given source of light is placed opposite a screen which is opaque to light and which has one or more apertures of definite shapes at definite points. We enquire what is the intensity of light at any point A behind the screen.

We can simplify the problem in the first place by restricting our attention to a single point C of the light-source. For the intensities of light which are due to the remaining points have merely to be added to the intensity which is due to C, in accordance with the results of § 36. We can imagine the wave-function which starts out from the point of light C as a Fourier series and restrict the whole calculation to a single term of the series. For in the case of natural light of constant intensity the individual terms of the series are independent of one another and lead to no appreciable interference effects.

To be able to apply Huygen's principle to the present case we place the Huygen plane in such a way that it completely shuts off both the source of light C and also the screen from the "point of diffraction" A, namely closely at the surface of the screen and in front of the opening ρ, on the side facing the point A. According to (157) and (158) we can then calculate the value of ϕ at the point A by integrating over the Huygen surface.

But, in order to be able to proceed, we now have to make a certain sacrifice in accuracy. The conditions are exactly the same as those encountered in applying Green's theorem (end of § 39). For, to introduce the quantity Ω in (158) into the calculation, we must know the values of ϕ and all its derivatives with respect to ν and t at all points of the Huygen surface; and this assumes that the problem of calculating ϕ is already

partially solved. This difficulty is accentuated by the fact that the values of ϕ, $\frac{\partial \phi}{\partial \nu}$ and $\frac{\partial \phi}{\partial t}$ at a definite point of the surface are not independent of one another, and that in general we cannot avoid arriving at an inherent contradiction if we introduce any approximate values into the expression for Ω. Of course, it is always possible to establish afterwards whether such a contradiction presents itself and what its nature is, by making the diffraction point A move into the Huygen plane and comparing the values calculated for ϕ and its derivatives from (157) with the values which were originally inserted.

In spite of this we shall here be able to proceed along the lines proposed. For innumerable applications have shown that Huygen's principle, not only when formulated rigorously but also when certain only approximately correct values of Ω are used, leads to results which are fully sufficient for practical optics. These approximate values of Ω are simply obtained by assuming the value of the expression for ϕ and its derivatives at all apertures of the screen to be the same as it would be if the screen were not present at all, and by assuming the value zero at all other points of the surface σ. The integral (157) is then only to be taken over the apertures of the screen, and in (158) we can ignore the screen altogether.

Of course, (157) resolves into just as many component integrals as there are apertures, and each component integral refers to a particular aperture. Moreover, it is easy to see that the value of the integral depends only on the contour of the aperture. For there is nothing to prevent the Huygen surface σ from assuming different positions which have the same contour. Hence besides speaking of a "diffracting aperture" we may also speak of a "diffracting contour or edge."

§ 44. In view of the above we are now concerned only with calculating the wave-function ϕ and its derivatives for all points of an aperture, without paying attention to the screen; a simple integration of the wave-equation

(146) serves for this purpose. We use as the simplest form for a particular solution of this equation, which corresponds to an absolutely homogeneous wave which starts out from the point-source of light C, by II (230), the expression :

$$\phi = \frac{A}{r_1} \cdot e^{i\omega\left(t - \frac{r_1}{c}\right)} \quad . \quad . \quad . \quad . \quad (161)$$

where r_1 denotes the distance of the reference point from the source S, to distinguish it from r, the distance of the diffraction point from the surface-element $d\sigma$. A is any complex constant, the other terms have the same meaning as before. Cf., for example, (83).

If we substitute the expression for ϕ given by (161) in (158) and perform the various differentiations, the quantity Ω_ν resolves into a number of terms which are of quite different orders of magnitude. For, since in the case of optical processes the wave-length $\lambda = \frac{2\pi c}{\omega}$ is small compared with the distances r and r_1 of the diffraction point A and the point-source C from the diffracting aperture, all those terms in Ω_ν which contain ω as a factor are large compared with the rest, so that the latter can be neglected (cf. III, § 88).

Finally, if we write $\tau = t - \frac{r}{c}$ in place of t, we get :

$$\Omega_\nu = \frac{2\pi i A}{\lambda r r_1}\left(\frac{\partial r_1}{\partial \nu} - \frac{\partial r}{\partial \nu}\right) e^{i\omega\left(t - \frac{r + r_1}{c}\right)}$$

and from (157) :

$$\phi_t(0) = \frac{iA}{2\lambda}\int \frac{d\sigma}{r r_1}\left(\frac{\partial r_1}{\partial \nu} - \frac{\partial r}{\partial \nu}\right) e^{i\omega\left(t - \frac{r + r_1}{c}\right)} \quad . \quad (162)$$

We shall simplify the further treatment of this problem by introducing a specialization which does not affect the characteristic features of diffraction phenomena. We shall first assume that the diffraction screen is plane, and secondly, that the light is incident normally, in parallel rays and wave-planes; that is, that r_1 has a constant value at all points of the diffracting apertures,

this value being great compared with the other dimensions involved. Then $\frac{\partial r_1}{\partial \nu} = 1$, whereas, since the direction of increasing ν forms an obtuse angle with that of increasing r, we may set $\frac{\partial r}{\partial \nu}$ equal to $-\cos\theta$, where θ denotes the "diffraction angle."

Hence (162) becomes:

$$\phi_t(0) = \frac{iA}{r_1} e^{i\omega\left(t-\frac{r_1}{c}\right)} \cdot \int \frac{d\sigma}{\lambda r} \cos^2\frac{\theta}{2} \cos e^{\frac{-2\pi i r}{\lambda}}$$

or:

$$\phi_t(0) = \frac{iA}{r_1} e^{i\omega\left(t-\frac{r_1}{c}\right)} (C - iS) \quad . \quad . \quad (163)$$

where we have used the abbreviations:

$$\left.\begin{aligned} C &= \frac{1}{\lambda}\int \frac{d\sigma}{r} \cos^2\frac{\theta}{2} \cos\frac{2\pi r}{\lambda} \\ S &= \frac{1}{\lambda}\int \frac{d\sigma}{r} \cos^2\frac{\theta}{2} \sin\frac{2\pi r}{\lambda} \end{aligned}\right\} \quad . \quad . \quad (164)$$

The real part of (163) represents a periodic vibration in time, a measure of whose intensity may be obtained, according to § 18, by multiplying the complex quantity (163) with its conjugate imaginary. Since, however, we are not concerned with the absolute value of the intensity of the light, but only with the ratio of the diffracted light to the incident light, we divide the quantity so found by the corresponding quantity for the incident light, namely by the product of the complex expression (161) and its conjugate imaginary: $\frac{|A|^2}{r_1^{\,2}}$.

In this way we arrive at the following expression for the diffracted light at the diffraction point A:

$$J = C^2 + S^2 \quad . \quad . \quad . \quad . \quad . \quad (165)$$

This reduces the diffraction problem to calculating the two integrals C and S, in which $d\sigma$ denotes any surface-element of the diffracting aperture, r its distance from the diffraction point A, and θ the angle of diffraction.

§ 45. The most characteristic peculiarity in the expressions for C and S is the different manner in which the quantities contained in them depend on the integration variables. For whereas θ and r change comparatively slowly with $d\sigma$, the trigonometrical functions, on account of the great value of $\frac{r}{\lambda}$, exhibit a very rapid fluctuation to and fro between a maximum and a minimum, with the sole exception of those surface-elements $d\sigma$ for which the distance r from the reference point is a minimum, because then r changes only slightly with $d\sigma$. This is the case for all those surface-elements $d\sigma$ which lie in the direct neighbourhood of the straight line which connects the point of light C with the diffraction point A and which is perpendicular to the plane of the screen.

Hence this straight line AC, which we may call the "central line," and in particular its point of intersection P with the Huygen plane, plays a characteristic part in determining the value of the intensity of the light at A.

But there is a special case for which the central line does not exist at all, namely, when the point A is situated at infinity, as well as C. Then AC denotes no particular straight line, but only a definite direction; no point of the Huygen plane plays a particular part and the integrals C and S change their character entirely. This is the case to which the diffraction phenomena discovered by and named after Fraunhofer belong.

§ 46. We shall first deal with the general case, the so-called *Fresnel diffraction phenomena*, starting with the question of the diffraction by a plane screen which is bounded by a single straight edge. The answer to this question is also of interest for the general case of any arbitrary edge, that is, for the general problem of shadow boundaries, because an edge of any shape whatsoever, provided that its curvature does not compare with the order of magnitude of the light-waves, may be imagined to consist solely of parts of straight lines.

We take as our plane of representation in Fig. 12 the plane through the diffraction point A and perpendicular to the edge of the screen; further, the line of intersection of the figure with the screen is the x-axis, the origin O in the edge of the screen, a plane through the edge of the screen and perpendicular to the plane of the figure is the y-axis, and the z-axis lies in the plane of the figure as shown. We designate the co-ordinates of the diffraction point A by $x, y = 0, z$; that of a point in the plane of the screen by $\xi, \eta, \zeta = 0$. If the aperture of the screen lies on the positive side of the x-axis, the integration is to be performed over all the surface-elements $d\sigma = d\xi,\ d\eta$ of the xy-plane with positive values of ξ, that is, in the case of ξ from 0 to ∞ and in the case of η from $-\infty$ to $+\infty$.

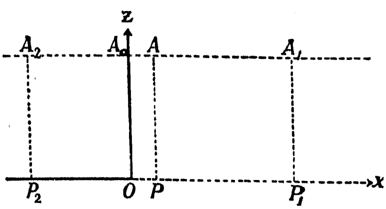

FIG. 12.

To calculate the integrals (164) we refer back to the above remarks on their nature and imagine the central line AC to be drawn, which cuts the screen-plane in the point P. We must now resolve the surface-integrals (164), which are to be performed over ξ and η, into two parts, namely firstly an integral over those points ξ, η which lie in the neighbourhood of the point P, so that their distance from P is small compared with $AP = z$, and secondly, an integral over all the remaining points of the region of integration. We shall reserve a more definite delimitation of the first component integral till later. After what has been said above we can replace the slowly variable factors r and θ in this integral by z and 0 respectively, and so obtain :

$$\left.\begin{aligned} C &= \frac{1}{\lambda z}\iint d\xi\, d\eta \cos\frac{2\pi r}{\lambda} + \ldots \\ S &= \frac{1}{\lambda z}\iint d\xi\, d\eta \quad \mathrm{n}\frac{2\pi r}{\lambda} + \ldots \end{aligned}\right\} \quad . \quad (166)$$

where the second component integral is denoted by the dots. To perform the integration first for η we write :

$$r^2 = (\xi - x)^2 + (\eta - y)^2 + (\zeta - z)^2$$

or, since :

$$y = 0 \text{ and } \zeta = 0 :$$

$$r^2 = \rho^2 + \eta^2 \quad . \quad . \quad . \quad . \quad . \quad (167)$$

where :

$$\rho^2 = (\xi - x)^2 + z^2 \quad . \quad . \quad . \quad . \quad (168)$$

remains constant during the integration. We integrate over η from $-\eta'$ to $+\eta'$, where :

$$\eta' << z \quad . \quad . \quad . \quad . \quad . \quad . \quad (169)$$

Since according to our assumptions $\eta << \rho$, we may expand (167) in a power series :

$$r = \rho + \frac{1}{2}\frac{\eta^2}{\rho} + \ldots \quad . \quad . \quad . \quad (170)$$

and omit the remaining terms.

If we substitute this expression in (166) and write :

$$\cos\frac{2\pi r}{\lambda} = \cos\frac{2\pi\rho}{\lambda}\cos\frac{\pi\eta^2}{\lambda\rho} - \sin\frac{2\pi\rho}{\lambda}\sin\frac{\pi\eta^2}{\lambda\rho} \quad (171)$$

the problem reduces to calculating the following two integrals :

$$\int_{-\eta'}^{+\eta'} d\eta \cos\frac{\pi\eta^2}{\lambda\rho} \text{ and } \int_{-\eta'}^{+\eta'} d\eta \sin\frac{\pi\eta^2}{\lambda\rho}. \quad . \quad (172)$$

or, if we use the abbreviation :

$$\eta\sqrt{\frac{\pi}{\lambda\rho}} = u \quad . \quad . \quad . \quad . \quad . \quad (173)$$

we may write the integrals as :

$$\sqrt{\frac{\lambda\rho}{\pi}}\int_{-u'}^{+u'} \cos u^2 du \text{ and } \sqrt{\frac{\lambda\rho}{\pi}}\int_{-u'}^{+u'} \sin u^2 du \quad . \quad (174)$$

The quantity :

$$u' = \eta'\sqrt{\frac{\pi}{\lambda\rho}} \quad . \quad . \quad . \quad . \quad . \quad (175)$$

may be regarded as very great compared with unity, in spite of the condition (169), so that :

$$\eta' >> \sqrt{\lambda z} \quad . \quad . \quad . \quad . \quad . \quad (176)$$

and, in writing this, we are supplementing our above convention concerning the domain of integration of C and S, respectively.

The two integrals in (174) then run :

$$\int_{-\infty}^{+\infty} \cos u^2 du \text{ and} \int_{-\infty}^{+\infty} \sin u^2 du \quad . \quad . \quad (177)$$

These two "Fresnel Integrals" have, as we shall see in the next paragraph, a finite value—namely $\sqrt{\frac{\pi}{2}}$ in both cases. If we substitute this value in (174), both integrals (172) assume the value $\sqrt{\frac{\lambda\rho}{2}}$, for which, on account of (169), we may also write $\sqrt{\frac{\lambda z}{2}}$, and so we obtain from (166), in view of (171), the expressions :

$$C = \frac{1}{\sqrt{2}}(c - s) + \ . \ . \ . \ , S = \frac{1}{\sqrt{2}}(c + s) + \ . \ . \ .$$

where we have used the abbreviations :

$$\left. \begin{aligned} c &= \frac{1}{\lambda z}\int \cos \frac{2\pi\rho}{\lambda}\, d\xi \\ s &= \frac{1}{\sqrt{\lambda z}}\int \sin \frac{2\pi\rho}{\lambda}\, d\xi \end{aligned} \right\} \quad . \quad . \quad . \quad . \quad (178)$$

and, finally, from (165) :

$$J = c^2 + s^2 + \ . \ . \ . \quad . \quad . \quad . \quad (179)$$

This reduces the problem to calculating the line-integrals c and s in (178), whose similarity to the integrals C and S in (166) is immediately obvious. Of course c and s, like C and S, are pure numbers.

A particularly striking feature of Fresnel's integrals

(177) is the fact that these integrals have a finite value although they stretch to infinity in both directions and although the functions that are to be integrated do not vanish at infinity, as is usually the case with such integrals, but retain finite values by fluctuating to and fro. The finite value of the integral results in spite of this because the maxima and minima $+1$ and -1 approach continually nearer each other as u increases, so that the fluctuations in the value of the integral become smaller and smaller. This peculiar behaviour is the mathematical counterpart of the physical circumstance that the contribution to the wave-function at the diffraction point A made by a strip $d\eta$ of the Huygen surface does indeed retain its order of magnitude, but changes its sign the more often the further the strip of surface is from the central line. Now what holds for the ineffective part played by the surface-elements $d\sigma$ whose co-ordinate η is of the order of magnitude η' in (176), must apply in a still greater measure to those surface-elements which are still further removed from the central line—that is, for all those surface-elements whose co-ordinates η are still greater than η'. Hence it follows that the second component integral (164) of C, whose limits lie between $\eta = \eta'$ and $\eta = \infty$, and also that of S, can make no appreciable contribution to these quantities. The fact that the factors r and θ in front of the cosine and sine can change appreciably does not come into question here. For their range of variation with η is so small comparatively that they can be neglected entirely for any interval of η, which comprises *many* maxima and minima of the cosines and sines.

Hence we may now omit the dots that were added in our formula and may also, in dealing with the Fresnel diffraction phenomena in the sequel, restrict the integration over the Huygen plane to such elements $d\sigma$ as lie near the central line, as determined by the conditions (169) and (176). From now onwards the following relation holds :

$$J = c^2 + s^2 \quad . \quad . \quad . \quad . \quad (179a)$$

§ 47. We shall make a digression here by discussing Fresnel's integrals. We have first to fill in a gap by calculating the definite integrals (177). For this purpose we start from Laplace's integral :

$$L = \int_{-\infty}^{+\infty} e^{-\alpha x^2} dx, \text{ where } \alpha > 0 \quad . \quad . \quad (180)$$

We may write it in the form :

$$L = \int_{-\infty}^{+\infty} e^{-\alpha y^2} dy.$$

Multiplying the two integrals together, we get :

$$L^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dxdy \, . \, e^{-\alpha(x^2+y^2)}$$

which, if we introduce polar co-ordinates :

$$dxdy = \rho \, . \, d\rho \, . \, d\phi$$

becomes transformed into :

$$L^2 = \int_0^{\infty} \int_0^{2\pi} \rho d\rho d\phi \, . \, e^{-\alpha\rho^2}$$

which works out to :

$$L^2 = \frac{\pi}{\alpha}$$

Hence :

$$L = \sqrt{\frac{\pi}{\alpha}}. \quad . \quad . \quad . \quad . \quad . \quad (181)$$

Fresnel's integrals are related to the Laplace integral in that i replaces α in the latter. For the sake of brevity we shall forgo the proof that the equation $\alpha = i$ also remains valid. Assuming this, we obtain from (180) :

$$L = \int_{-\infty}^{+\infty} (\cos x^2 - i \sin x^2) \, dx$$

and from (181) :

$$L = \sqrt{\frac{\pi}{i}} = (1 - i) \, . \sqrt{\frac{\pi}{2}}$$

If we equate the last two expressions we arrive at the values already used in the preceding section :

$$\left.\begin{aligned}\int_{-\infty}^{+\infty} \cos u^2 du &= \sqrt{\frac{\pi}{2}}\\ \int_{-\infty}^{+\infty} \sin u^2 du &= \sqrt{\frac{\pi}{2}}\end{aligned}\right\} \quad . \quad . \quad . \quad . \qquad (182)$$

Since we also require the values of Fresnel's integrals for variable limits, we shall carry the investigation a little further. Since $(+u)^2 = (-u)^2$ we easily obtain the integral with the limit 0 :

$$\int_{-\infty}^{0} \cos u^2 du = \int_{0}^{\infty} \cos u^2 du = \int_{0}^{\infty} \sin u^2 du = \frac{1}{2}\sqrt{\frac{\pi}{2}}. \qquad (183)$$

On the other hand, it is not possible to reduce the integral with one variable limit u to terms of elementary functions. We are compelled to resort to expansions in series.

An expansion which is valid for the integral taken from 0 to u and which is always convergent is derived by applying the formulæ of integration by parts :

$$\int_0^u \cos u^2 du = u \cos u^2 + 2\int_0^u u^2 \sin u^2 du$$

$$\int_0^u \sin u^2 du = u \sin u^2 - 2\int_0^u u^2 \cos u^2 du$$

If we repeat the operation of integration by parts on the integrals on the right, we ultimately get :

$$\int_0^u \cos u^2 du = \cos u^2 . \Sigma + \sin u^2 . \Gamma \quad . \quad . \qquad (184)$$

$$\int_0^u \sin u^2 du = \sin u^2 . \Sigma - \cos u^2 . \Gamma \quad . \quad . \qquad (185)$$

where we have used the abbreviations :

$$\Sigma = u - \frac{2}{3}\cdot\frac{2}{5}u^5 + \frac{2 . 2 . 2 . 2}{3 . 5 . 7 . 9}u^9 - \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (186)$$

$$\Gamma = \frac{2}{3}u^3 - \frac{2 . 2 . 2}{3 . 5 . 7}u^7 + \frac{2 . 2 . 2 . 2 . 2}{3 . 5 . 7 . 9 . 11}u^{11} - \quad . \quad . \quad . \qquad (187)$$

These series converge for all positive and negative values

of u, because the quotient of two successive terms decreases to an unlimited extent as the order of the term increases. But on account of the rapid increase of the power indices they are quite unsuitable for calculation for u's of any considerable magnitude.

In the case of greater values of u we find it more advantageous to use expansions in decreasing powers of u, expansions which are often divergent but, as we shall see, are sometimes of great service in the calculations. Since a series with negative powers of u approximates to the value zero as u increases, it is expedient to start out from the integral which has $u = \infty$ as its limit. We shall therefore now apply integration by parts in the following form:

$$\int_{u>0}^{\infty} \cos u^2 du = \int_u^{\infty} u \cos u^2 \cdot \frac{1}{u} du = -\frac{\sin u^2}{2u} + \int_u^{\infty} \frac{\sin u^2 du}{2u^2}$$

$$\int_u^{\infty} \sin u^2 du = \int_u^{\infty} u \sin u^2 \cdot \frac{1}{u} du = \frac{\cos u^2}{2u} - \int_u^{\infty} \frac{\cos u^2 du}{2u^2}$$

and we obtain by applying the same process to the integrals on the right-hand side:

$$\int_u^{\infty} \cos u^2 du = \cos u^2 . \gamma_n - \sin u^2 . \sigma_n + R_c \quad . \quad (188)$$

$$\int_u^{\infty} \sin u^2 du = \sin u^2 . \gamma_n + \cos u^2 . \sigma_n + R_s \quad . \quad (189)$$

The series:

$$2\sigma_n = \frac{1}{u} - \frac{1.3}{2.2} \cdot \frac{1}{u^5} + \ldots (-1)^{n-1} . \frac{1.3 \ldots (4n-5)}{2.2 \ldots \quad 2} \cdot \frac{1}{u^{4n-3}} \quad . \quad . \quad (190)$$

$$2\gamma_n = \frac{1}{2} \cdot \frac{1}{u^3} - \frac{1.3.5}{2.2.2} \cdot \frac{1}{u^7} + \ldots (-1)^{n-1} . \frac{1.3 \ldots (4n-3)}{2.2 \ldots \quad 2} \cdot \frac{1}{u^{4n-1}} \quad . \quad . \quad (191)$$

are divergent and hence have been broken off at a definite order number n.

The remainder terms are :

$$R_c = (-1)^n \cdot \frac{1 \cdot 3 \ldots (4n-1)}{2 \cdot 2 \ldots 2} \cdot \int_u^\infty \frac{\cos u^2 du}{u^{4n}} \quad . \quad (192)$$

$$R_s = (-1)^n \cdot \frac{1 \cdot 3 \ldots (4n-1)}{2 \cdot 2 \ldots 2} \cdot \int_u^\infty \frac{\sin u^2 du}{u^{4n}} \quad . \quad (193)$$

They both increase beyond all limits as n increases. Nevertheless these series may often be used successfully in calculations. For closer inspection shows that for certain values of the order number n the remainder terms become very small and hence can be entirely neglected without introducing an appreciable error.

For in the first place it is easy to see that both R_c and R_s are less than the expression R_n which is obtained if we replace the cosine and sine in the integrals (192) and (193) respectively, by 1, that is less than :

$$R_n = \frac{1 \cdot 3 \ldots (4n-1)}{2 \cdot 2 \ldots 2} \int_u^\infty \frac{du}{u^{4n}}$$

$$= \frac{1 \cdot 3 \ldots (4n-3)}{2 \cdot 2 \ldots 2} \cdot \frac{1}{2} \cdot \frac{1}{u^{4n-1}} \quad . \quad (194)$$

This is precisely the last term in the series for γ_n, a number which, in certain circumstances, is extremely small, particularly when u is great. The essential step is, of course, the choice of the order number n. If we allow n to increase from 1 onwards, R_n first decreases, on account of the power $(4n-1)$ in the denominator, but afterwards increases to an unlimited extent owing to the factorial term in the numerator.

The minimum value of R_n is that at the order number n, for which the ratio $R_{n+1} : R$ approaches most nearly to 1. This is the case when :

$$\frac{R_{n+1}}{R_n} = \frac{(4n-1)}{2} \cdot \frac{(4n+1)}{2} \cdot \frac{1}{u^4} \text{ nearly} = 1.$$

$$\text{That is, } n \text{ nearly} = \frac{u^2}{2} \quad . \quad . \quad . \quad (195)$$

The greater the value of u, the more the number of terms that must be given to the series γ_n and σ_n in order

that the remainder terms shall become as small as possible and the smaller these remainder terms become. If we neglect them we obtain approximate values for Fresnel's integrals from (188) and (189). Such series, which, although divergent, approach a definite limit where a definite number of terms are taken but afterwards recede from this limit again, are said to be "semi-convergent." They can often be used to better purpose than convergent series, but have the fundamental disadvantage that the approximation cannot be carried as far as we please, as in the case of convergent series, but comes to an end at a definite point.

Nevertheless, even for moderate values of u the semi-convergent series γ_n and σ_n are much more convenient than the convergent series Γ and Σ for determining the approximate value of Fresnel's integrals. For $u = 2$, for example, there are already very many terms to be taken into account in the convergent series (186) and (187) if the fluctuations in the sum are to be made even moderately small, whereas the remainder term R of the divergent expansion, in which, by (195), n has been put equal to 2, has, by (194), the value :

$$R_2 = \frac{1 \,.\, 3 \,.\, 5}{2 \,.\, 2 \,.\, 2} \cdot \frac{1}{2} \cdot \frac{1}{2^7} = 0{\cdot}0073$$

which, compared with 1, is already a fairly small number and can in most circumstances be neglected without danger.

§ 48. Reverting to our problem of determining the intensity of light J at the diffraction point A, we proceed to calculate the decisive integrals c and s in (178), by first expressing ρ in terms of the integration variable ξ in accordance with (168). Since, as we showed in § 46, we need take into consideration only the points near the central line, we may assume $\xi - x$ to be small compared with z and write the following approximate value for ρ :

$$\rho = z + \frac{(\xi - x)^2}{2z} \quad . \quad . \quad . \quad . \quad . \quad (196)$$

Thus :

$$\cos\frac{2\pi\rho}{\lambda} = \cos\frac{2\pi z}{\lambda}\cos\frac{\pi(\xi - x)^2}{\lambda z} - \sin\frac{2\pi z}{\lambda}\sin\frac{\pi(\xi - x)^2}{\lambda z}.$$

If we substitute this in the expression (178) for c, we obtain the two integrals in it as :

$$\int \cos\frac{\pi(\xi - x)^2}{\lambda z}\cdot d\xi \text{ and } \int \sin\frac{\pi(\xi - x)^2}{\lambda z}\cdot d\xi$$

which are both to be taken between the limits $\xi = 0$ and $\xi = \infty$. By introducing the integration variables :

$$u = (x - \xi)\cdot\sqrt{\frac{\pi}{\lambda z}} \quad . \quad . \quad . \quad . \quad (197)$$

we may replace them by the integrals :

$$\sqrt{\frac{\lambda z}{\pi}}\int_{-\infty}^{u}\cos u^2 du \text{ and } \sqrt{\frac{\lambda z}{\pi}}\int_{-\infty}^{u}\sin u^2 du$$

where u, the upper limit, denotes the value of (197) for $\xi = 0$, that is :

$$u = \sqrt{\frac{\pi}{\lambda z}}\cdot x \quad . \quad . \quad . \quad . \quad . \quad (198)$$

This transforms the expression (178) for c into :

$$c = \frac{1}{\sqrt{\pi}}\left(\cos\frac{2\pi z}{\lambda}\cdot C - \sin\frac{2\pi z}{\lambda}\cdot S\right)$$

where we have now used the abbreviations :

$$C = \int_{-\infty}^{u}\cos u^2 du \text{ and } S = \int_{-\infty}^{u}\sin u^2 du \quad . \quad (199)$$

And in the same way we get from (178) :

$$s = \frac{1}{\sqrt{\pi}}\left(\sin\frac{2\pi z}{\lambda}\cdot C + \cos\frac{2\pi z}{\lambda}\cdot S\right)$$

from which, by (179*a*), we get for the desired value of the intensity of light :

$$J = \frac{1}{\pi}(C^2 + S^2) \quad . \quad . \quad . \quad . \quad (200)$$

This reduces the whole problem to the calculation of the Fresnel integrals C and S, in which u has the value given in (198).

Since the intensity of light J depends only on u, it is quite independent of y, as is obvious, and we can restrict our attention to the plane of Fig. 12, $y = 0$. Moreover, on every curve :

$$\frac{x}{\sqrt{z}} = \text{const.}$$

we have that the intensity of the light is constant. This is a branch of a parabola which touches the x-axis at its end and vertex O and is more or less strongly curved at O. If we make the constant assume all values from $-\infty$ to $+\infty$ and bear in mind that $\sqrt{z}$ is always positive, but x, the abscissa of the diffraction point A, is either positive or negative, we first get $z = 0, x < 0$—that is, the points which lie immediately behind the screen, and then the parabolic branch which lies on the side of the screen; further, for const. $= 0$ we have the positive z-axis, and then the branches which lie on the side of the aperture, as far as the positive x-axis : $z = 0, x > 0$.

To obtain a survey of the relation between the intensity of the light J on the position of the point A, it is therefore sufficient to displace the diffraction point A along any line $z =$ const. parallel to the x-axis. By (198) u is then a direct measure of the abscissa x of the diffraction point or of the distance OP of the central line AP from the edge of the screen.

We shall now allow the diffraction point A to move along such a line (Fig. 12) from the right to the left across the whole plane of the figure, and shall inquire as to the intensity at such point as given by (200) and (199).

1. For very great positive values of u (for which A is very far to the right, say at A_1) we can set the upper limits of the integrals (199) equal to ∞, and we obtain from (182) and (200) that $J = 1$—that is, at a great distance from the screen the intensity of the light is

just as great as if the screen were not present, as is to be anticipated.

2. If u has moderate positive values, then, as we have seen, the expansions in semi-convergent series give us a useful approximation for estimating the values of Fresnel's integrals. In this case we can even go down as far as about $u = 1{\cdot}5$. For then, when $n = 1$, the remainder terms R_c and R_s which are to be neglected are, by (194) always smaller than :

$$R_1 = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{1{\cdot}5^3} = 0{\cdot}074.$$

To be able to apply the formulæ (188) and (189) we write instead of (199) :

$$C = \int_{-\infty}^{+\infty} \cos u^2 du - \int_{u}^{+\infty} \cos u^2 du$$

$$S = \int_{-\infty}^{+\infty} \sin u^2 du - \int_{u}^{+\infty} \sin u^2 du.$$

On account of (182) we obtain the approximate values :

$$\left.\begin{aligned} C &= \sqrt{\frac{\pi}{2}} - \gamma_n \cos u^2 + \sigma_n \sin u^2 \\ S &= \sqrt{\frac{\pi}{2}} - \gamma_n \sin u^2 - \sigma_n \cos u^2 \end{aligned}\right\} \quad . \quad . \quad (201)$$

Consequently, by (200), we get :

$$J = \frac{1}{\pi}\Big[\pi - 2\sqrt{\pi}\gamma_n \cos\left(u^2 - \frac{\pi}{4}\right) + 2\sqrt{\pi}\sigma_n \sin\left(u^2 - \frac{\pi}{4}\right) + \gamma_n^2 + \sigma_n^2\Big] \quad . \quad . \quad (202)$$

The question arises whether the intensity of light J has maxima and minima in the region of u under consideration here. By (200) such points must result from the equation :

$$C\frac{dC}{du} + S\frac{dS}{du} = 0 \quad . \quad . \quad . \quad . \quad (203)$$

or, since by (199) :

$$\frac{dC}{du} = \cos u^2 \text{ and } \frac{dS}{du} = \sin u^2 \quad . \quad . \quad (204)$$

we have:

$$\cos\left(u^2 - \frac{\pi}{4}\right) = \frac{\gamma_n}{\sqrt{\pi}}$$

$$= \frac{1}{2\sqrt{\pi}}\left(\frac{1}{2}\cdot\frac{1}{u^3} - \frac{1}{2}\cdot\frac{3}{2}\cdot\frac{5}{2}\cdot\frac{1}{u^7} + \ . \ . \ .\right). \qquad (205)$$

The expression on the right-hand side is very small for all the values of u that come into question; consequently the equation will be satisfied by such values of the argument $u^2 - \frac{\pi}{4}$ as approximate very closely to one of the values $\frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, \frac{7\pi}{2}, \ . \ . \ .$ and this gives the following approximate values of u:

$$u_1 = \sqrt{\frac{3\pi}{4}},\ u_2 = \sqrt{\frac{7\pi}{4}},\ u_3 = \sqrt{\frac{11\pi}{4}},\ u_4 = \sqrt{\frac{15\pi}{4}},\ . \ . \ . \qquad (206)$$

The first and smallest of these values u_1 is equal to 1·535 and lies just within the range of u which we are considering.

In this range we thus actually have an infinite number of maxima and minima of intensity. We determine whether a value is a maximum or a minimum by substituting in (202). Since γ_n and σ_n are small positive numbers, the value of J is essentially influenced by the term with $\sigma_n \sin\left(u^2 - \frac{\pi}{4}\right)$ and the sign of the sine decides whether J is a maximum or a minimum, greater or less than 1. So we find that J is a maximum for the points u_1, u_3, u_5 . . . and a minimum for the points u_2, u_4, u_6, . . . From (202) and (190) we get as approximate values for the maxima and minima:

$$J = 1 \pm \frac{1}{u\sqrt{\pi}}$$

the maxima being, by (206):

$$J_1 = 1 + \frac{2}{\pi\sqrt{3}},\ J_3 = 1 + \frac{2}{\pi\sqrt{11}},\ . \ . \ .$$

and the minima :

$$J_2 = 1 - \frac{2}{\pi\sqrt{7}}, \; J_4 = 1 - \frac{2}{\pi\sqrt{15}}, \; . \quad .$$

As u increases both these points and the values of the maxima and minima continually approach each other until they finally blur each other.

3. If the diffraction point A approaches so near to the z-axis or the shadow boundary of the screen that the abscissa u has rather small positive or negative values, say between $-1{\cdot}5$ and $+1{\cdot}5$, we find it advantageous, in calculating the Fresnel integrals, to use the expansions in convergent series. We therefore write (199) in the form :

$$C = \int_{-\infty}^{0} \cos u^2 du + \int_{0}^{u} \cos u^2 du$$

$$S = \int_{-\infty}^{0} \sin u^2 du + \int_{0}^{u} \sin u^2 du$$

and by (183), (184), (185) and (200) :

$$J = \frac{1}{\pi}\left[\frac{\pi}{4} + \sqrt{\pi} \, . \, \Sigma \, . \cos\left(u^2 - \frac{\pi}{4}\right) \right.$$
$$\left. + \sqrt{\pi} \, . \, \Gamma \, . \sin\left(u^2 - \frac{\pi}{4}\right) + \Sigma^2 + \Gamma^2\right] \quad . \qquad (207)$$

If, to find whether there are maxima and minima of J in this region we again proceed as above, we arrive at the condition :

$$\frac{\sqrt{\pi}}{2} \cos\left(u^2 - \frac{\pi}{4}\right) + \Sigma = 0$$

or, by (186) :

$$\cos\left(u^2 - \frac{\pi}{4}\right) + \frac{2}{\sqrt{\pi}}\left(u - \frac{2}{3} \cdot \frac{2}{5} u^5 + \; . \; . \; .\right) = 0 \; . \qquad (208)$$

In this equation the first summand is always positive within the region under consideration $(-1{\cdot}5 < u < +1{\cdot}5)$; the second changes its sign with u, but its absolute value never reaches that of the cosines. To see this it is sufficient to calculate the series Σ for the extreme

case $u = 1{\cdot}5$. We can then find the degree of approximation which corresponds to the number of terms used by narrowing the neglected remainder term to the extent corresponding to (194).

The question of the roots of the equation $\frac{dJ}{du} = 0$ may be illustrated still more strikingly if we represent geometrically the dependence of the function J on u; this dependence is fully fixed by the simple equations (200) and (204) taken in conjunction with the boundary conditions for $u = 0$. This is done by taking C and S as rectangular co-ordinates of a point in a plane and interpreting J as the square of the distance of this point from the origin.

In accordance with what has been stated above, the expression (208) has the sign of the cosine in the whole region of u under consideration—that is, it is positive, or $\frac{dJ}{du} > 0$, and the intensity of the light increases steadily as the diffraction point A passes from the negative to the positive abscissæ, until it reaches the first maximum at u_1. In the immediate vicinity of the point A_0 of the boundary of the geometric shadow we have by (207) and (186):

$$J = \frac{1}{4} + \frac{u}{\sqrt{2\pi}} \quad . \quad . \quad . \quad . \quad . \quad (209)$$

In A_0 itself we have $J = \frac{1}{4}$.

4. If, lastly, the diffraction point A moves a greater distance from the shadow-limit ($u < -1{\cdot}5$) towards the left, say to A_2, we can again fall back on our semi-convergent series by writing:

$$u = -u'(u' > 1{\cdot}5)$$

$$C = \int_{-\infty}^{-u'} \cos u^2 du = \int_{u'}^{\infty} \cos u^2 du$$

$$S = \int_{-\infty}^{-u'} \sin u^2 du = \int_{u'}^{\infty} \sin u^2 du$$

and by using the formulæ (188) and (189) for these expressions, omitting the remainder terms, this gives :

$$J = \frac{1}{\pi}(\gamma_n^2 + \sigma_n^2) \quad . \quad . \quad . \quad . \quad (210)$$

whereas the condition for a maximum or a minimum reduces, by (203) and (204), to $\gamma_n = 0$, which is fulfilled nowhere. Hence the intensity of the light decreases steadily as u' increases. For very great values of u' we finally get :

$$J = \frac{1}{4\pi u^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (211)$$

and for $u = -\infty$ we get $J = 0$.

The above results also fix the intensity J for all other points on the xz-plane. For a definite parabolic branch of the family $\frac{x}{\sqrt{z}} = \text{const.}$ passes through every point A of the straight line $A_2 A_0 A_1$ in question, and the intensity of light at A is the same at all points of any particular parabolic arc as far as the singular point O, where all the curves touch. Thus when the straight line A_2A_1 approaches the x-axis, always remaining parallel to it, the maxima and the minima of J draw closer and closer together, without altering their values, and when the straight line coincides with the x-axis, they all concentrate in O, while the negative axis of x represents the parabolic branch of intensity zero and the positive axis of x represents that of intensity unity. This accords fully with the assumptions with which we started in § 43 in applying Huygen's principle.

§ 49. Having treated the general case of Fresnel's diffraction phenomena, we shall now link up with the remarks made at the conclusion of § 45 and discuss the special case of *Fraunhofer's diffraction phenomena*, which have incomparably greater significance for practical optics and which are fortunately much more amenable to mathematical treatment. We shall again start from the equations (164) and (165), which hold for a plane diffraction screen on which light falls normally. A further special

assumption is that not only the point-source C but also the diffraction point A lie at infinity. We shall use the same notation as earlier—that is, we shall choose the plane of the screen as the xy-plane and shall take x, y, z as the co-ordinates of the diffraction point, and ξ, η, $\zeta = 0$ as the co-ordinates of a point of the aperture. We shall take as our origin any point within the aperture.

Since the distance of the diffraction point A from all points of the diffracting aperture is to be infinitely great compared with the linear dimensions of the aperture, we shall assume the latter to be entirely in finite regions. Then the rays which travel from points in the aperture to A are all parallel and their direction is characteristic of the position of A. For the distance of A plays no part physically. Hence in this case it is sufficient to speak of the *direction of diffraction* instead of the diffraction point. This direction is defined by the two polar angles θ and ϕ in the equations :

$$\frac{x}{r_0} = \sin\theta\cos\phi, \quad \frac{y}{r_0} = \sin\theta\sin\phi, \quad \frac{z}{r_0} = \cos\theta \quad . \qquad (212)$$

where x, y, z and r_0, the distance OA, are infinitely great.

When substituting these values in (164) we must note that in the denominator r is to be regarded as constant, as well as θ, in the integration, while in the ratio $\frac{r}{\lambda}$ the variability of the numerator becomes of decisive importance. Thus :

$$\begin{aligned} r^2 &= (x-\xi)^2 + (y-\eta)^2 + (z-\zeta)^2 \\ &= r_0^2 - 2x\xi - 2y\eta \\ r &= r_0 - \frac{\xi x + \eta y}{r_0} \end{aligned}$$

If we substitute this value in (164), the expressions for C and S may be written in the following form :

$$C = \cos\frac{2\pi r_0}{\lambda}\cdot c + \sin\frac{2\pi r_0}{\lambda}\cdot s$$

$$S = \sin\frac{2\pi r_0}{\lambda}\cdot c - \cos\frac{2\pi r_0}{\lambda}\cdot s$$

where :

$$\left.\begin{aligned} c &= \frac{1}{\lambda r_0}\cos^2\frac{\theta}{2}\cdot\int d\sigma\cos(\alpha\xi+\beta\eta) \\ s &= \frac{1}{\lambda r_0}\cos^2\frac{\theta}{2}\cdot\int d\sigma\sin(\alpha\xi+\beta\eta) \end{aligned}\right\} \quad . \quad . \qquad (213)$$

$$\frac{2\pi x}{\lambda r_0} = \frac{2\pi\sin\theta\cos\phi}{\lambda} = \alpha, \quad \frac{2\pi y}{\lambda r_0} = \frac{2\pi\sin\theta\sin\phi}{\lambda} = \beta. \qquad (214)$$

This, combined with (165), gives us for the intensity J in the direction of diffraction (θ, ϕ) the value :

$$J = c^2 + s^2 \quad . \quad . \quad . \quad . \quad . \quad . \qquad (215)$$

Since α and β remain constant during the integration, there is no difficulty in performing it.

For the diffraction angle $\theta = 0$ (diffraction point A on the z-axis) $\alpha = \beta = 0$, thus $s = 0$ also, and we obtain the " axial " intensity of light from (215) and (213) as :

$$J_0 = \frac{\sigma^2}{\lambda^2 r_0^2} \quad . \quad . \quad . \quad . \quad . \quad . \qquad (216)$$

where σ denotes the area of the aperture. What strikes us at once as unusual about this formula is that it is proportional to σ^2 and not to σ; for the energy which passes from the source of light through the aperture is certainly proportional to σ. This apparent paradox is explained by the fact that J_0 does not directly denote the energy of radiation; rather, a finite amount of energy is radiated only within a finite cone from the directions of diffraction (θ, ϕ), as we established in § 18.

Since we are not concerned with absolute values in finding the intensity of the diffracted light, it is more convenient to reduce it to terms of the axial intensity and to write instead of (215) :

$$J = J_0 \cdot (C^2 + S^2) \quad . \quad . \quad . \quad . \quad . \qquad (217)$$

where, by (216) and (213), we have the newer meaning :

$$\left.\begin{aligned} C &= \cos^2\frac{\theta}{2}\cdot\frac{1}{\sigma}\int d\sigma\cos(\alpha\xi+\beta\eta) \\ S &= \cos^2\frac{\theta}{2}\cdot\frac{1}{\sigma}\int d\sigma\sin(\alpha\xi+\beta\eta). \end{aligned}\right\} \quad . \qquad (218)$$

§ 50. As an illustration let us calculate the diffraction by a *rectangular* aperture of length l and width b, so that $\sigma = bl$. If we take the centre of the rectangle as the origin O, we have in the integration :

$$-\frac{b}{2} < \xi < +\frac{b}{2}, \quad -\frac{l}{2} < \eta < +\frac{l}{2}$$

and we get from (218) :

$$C = \cos^2\frac{\theta}{2} \cdot \frac{\sin\frac{\alpha b}{2}}{\frac{\alpha b}{2}} \cdot \frac{\sin\frac{\beta l}{2}}{\frac{\beta l}{2}}$$

$$S = 0$$

and from (217) :

$$J = J_0 \,.\, \cos^4\frac{\theta}{2} \cdot \frac{\sin^2\frac{\alpha b}{2}}{\left(\frac{\alpha b}{2}\right)^2} \cdot \frac{\sin^2\frac{\beta l}{2}}{\left(\frac{\beta l}{2}\right)^2} \quad . \quad . \quad . \quad (219)$$

This, in conjunction with (214), is an explicit relation between the intensity J of the light and the direction of diffraction (θ, ϕ).

For $\theta = 0$ we get $J = J_0$, as should be. This axial intensity is the greatest that occurs, which agrees with the circumstance that the function $\frac{\sin x}{x}$ has its greatest value when $x = 0$. As x increases this function fluctuates to and fro between a constantly decreasing maximum and the minimum zero at regular intervals, until it ultimately vanishes. Moreover, J is doubly variable in that two systems of equi-distant zero-points become superposed. The one is denoted by the values :

$$\alpha = \frac{2\pi x}{\lambda r_0} = \frac{2\pi}{b}, \frac{4\pi}{b}, \frac{6\pi}{b}, \ldots$$

the other, by :

$$\beta = \frac{2\pi y}{\lambda r_0} = \frac{2\pi}{l}, \frac{4\pi}{l}, \frac{6\pi}{l}, \ldots \quad . \quad . \quad . \quad (220)$$

Between these two zero points there is a maximum which decreases as the order number increases. To this there is added the gradual decrease of J as the angle of diffraction θ increases.

These variations of intensity may be observed by allowing the diffracted light to fall on a condensing lens. A definite point in the focal plane of the lens then corresponds to all the rays diffracted in a definite direction (θ, ϕ) and our above discussion shows that the diffraction pattern that appears will be a bright cross with two systems of dark bands on both sides, situated symmetrically with respect to the edges of the rectangle. The narrower the breadth of the rectangular aperture and the longer the wave-length used, the more the bands are separated. If we have a long and narrow rectangle (slit), the bands of the one system merge together and only the other system, parallel to the length of the slit, remains.

This diffraction picture does not, of course, represent an image of the diffracting aperture in the sense of geometric optics; for we are not dealing with optically conjugate points here. Hence we must not expect a geometrical similarity between the diffraction image and the diffracting aperture.

§ 50*a*. The practical importance of Fraunhofer's diffraction phenomena becomes apparent when we use a great number of diffracting apertures congruent to one another. We assume a number N of such congruent apertures, of arbitrary form and arbitrary distribution but similarly placed on the screen. The intensity J of the diffracting light can then be reduced simply to the intensity J_1 of the diffracting light of a single aperture.

For we refer the co-ordinates ξ, η of the points of the nth aperture to any arbitrarily chosen point O_n of the aperture as origin, by writing :

$$\xi = \xi_n + \xi', \ \eta = \eta_n + \eta' \quad . \quad . \quad . \quad (221)$$

where ξ_n and η_n are the co-ordinates of O_n with respect to the general origin of co-ordinates O, which may be

chosen anywhere, and we take the points O_n in all the other apertures at the corresponding point. Then each of the two integrals (218) resolves into N separate integrals each of which is to be taken over one aperture, with the given limits ξ' and η', which are the same for each integral. We obtain :

$$C=\cos^2\frac{\theta}{2}\cdot\frac{1}{\sigma}\sum_{n=1}^{n=N}\iint d\xi' d\eta' \cos\{\alpha(\xi_n+\xi')+\beta(\eta_n+\eta')\} \quad . \qquad (222)$$

The summation refers to the index n, the integration to the co-ordinates ξ', η'. These two operations are quite independent of one another, and so can have their order reversed at will.

If, analogously to (218), we use the abbreviations :

$$\left.\begin{aligned}\cos^2\frac{\theta}{2}\cdot\frac{1}{\sigma_1}\iint d\xi' d\eta' \cos(\alpha\xi'+\beta\eta') &= C_1\\ \cos^2\frac{\theta}{2}\cdot\frac{1}{\sigma_1}\iint d\xi' d\eta' \sin(\alpha\xi'+\beta\eta') &= S_1\end{aligned}\right\} \quad . \qquad (223)$$

where, from now onwards, we use the index 1 to denote the quantities that refer to a single aperture, and likewise :

$$\sigma = N\,.\,\sigma_1 \quad . \quad . \quad . \quad . \quad . \quad . \qquad (224)$$

and :

$$\left.\begin{aligned}\sum_{n=1}^{n=N}\cos(\alpha\xi_n+\beta\eta_n) &= C_N\\ \sum_{n=1}^{n=N}\sin(\alpha\xi_n+\beta\eta_n) &= S_N\end{aligned}\right\} \quad . \quad . \quad . \quad . \qquad (225)$$

we get, by (222) :

$$C=\frac{1}{N}(C_N C_1 - S_N S_1)$$

Similarly :

$$S=\frac{1}{N}(S_N C_1 + C_N S_1)$$

and, by (217) :

$$J=\frac{J_0}{N^2}(C_1{}^2+S_1{}^2)(C_N{}^2+S_N{}^2).$$

Lastly, by (216) and (224):

$$J = J_{01}\,(C_1{}^2 + S_1{}^2)(C_N{}^2 + S_N{}^2) \quad . \quad . \quad (226)$$

for which we may now write, by (217):

$$J = J_1\,(C_N{}^2 + S_N{}^2) \quad . \quad . \quad . \quad . \quad (227)$$

This simple relationship reduces the intensity of the diffracted light of the system of apertures to terms of that of a single aperture. It shows that the use of *several* apertures does not cause an extension and broadening of the diffraction pattern in space, but only an increase in the intensity of light of the diffraction pattern due to a *single* aperture. But we shall see that the intensification factor assumes totally different values according as the apertures are distributed regularly or irregularly over the screen. For by (225) we have:

$$\begin{aligned} C_N{}^2 + S_N{}^2 &= \sum_{n=1}^{N} \cos^2(\alpha\xi_n + \beta\eta_n) \\ &\quad + 2\sum_{n=1}^{N-1}\sum_{n'=n+1}^{N} \cos(\alpha\xi_n + \beta\eta_n)\cos(\alpha\xi_{n'} + \beta\eta_{n'}) \\ &\quad + \sum_{n=1}^{N} \sin^2(\alpha\xi_n + \beta\eta_n) \\ &\quad + 2\sum_{n=1}^{N-1}\sum_{n'=n+1}^{N} \sin(\alpha\xi_n + \beta\eta_n)\sin(\alpha\xi_{n'} + \beta\eta_{n'}) \\ &= N + 2\sum_{n=1}^{N-1}\sum_{n'=n+1}^{N} \cos\{\alpha(\xi_{n'} - \xi_n) + \beta(\eta_{n'} - \eta_n)\} \quad (228) \end{aligned}$$

If the apertures are distributed quite irregularly—as, for example, in the case of a card which has been pierced by a pin indiscriminately, then the terms after the double summation sign have partly positive and partly negative values; this causes the sum to fluctuate violently for different values of α and β, but it acquires no finite mean value. The result is that, by (227), the intensity of the diffracted light amounts, except for local fluctuations, to N-times the intensity of the diffracted light for a single aperture.

The result that emerges is quite different if we use

regularly distributed apertures. If, for example, the points O_n lie so that for certain values of α and β all angles that occur in (228) are exact multiples of 2π, the double sum becomes equal to $N(N-1)$, and so, for great values of N, it becomes of the order N^2.

This peculiar behaviour is, of course, due to the fact that in the case of irregular distribution the rays that are diffracted in a certain direction by the N-apertures do not interfere appreciably with one another, so that their intensities simply become added together, whereas in the case of regular distribution these N-rays can all be in the same phase, so that the absolute values of the amplitudes of the field-strengths add themselves together, and not the intensities. Now since the resultant intensity is represented by the square of the resultant amplitude, it amounts to N^2 times the intensity due to a *single* aperture. It is on this enormous increase in the intensity of the light that the great practical importance of Fraunhofer's diffraction phenomena depends.

§ 51. We shall perform the calculation for the case of a *grating*.

Let b be the width of aperture of a single slit, c the grating-constant—that is, the distance between two corresponding points of two neighbouring slits or the sum of the width of a slit and of the space between two neighbouring slits. Then we may set all the η's in (225) equal to zero, whereas :

$$\xi_1 = 0,\ \xi_2 = c,\ \xi_3 = 2c,\ \ldots\ \xi_N = (N-1)c.$$

We may simplify the summation of the two trigonometric series (225) considerably by recollecting that the expression $C_N^2 + S_N^2$, with which we are alone concerned here, is no other than the square of the absolute value of the complex quantity :

$$C_N + iS_N = \sum_{n=1}^{N} e^{i\alpha\xi_n} = \sum_{n=1}^{N} e^{i\alpha(n-1)c} = \frac{e^{i\alpha Nc} - 1}{e^{i\alpha c} - 1}.$$

By multiplying this quantity with its conjugate imagin-

ary we obtain the desired quantity and from this, by (227), the intensity of the diffracted light :

$$J = J_1 \frac{\sin^2 N \frac{\alpha c}{2}}{\sin^2 \frac{\alpha c}{2}} \quad . \quad . \quad . \quad . \quad . \quad (229)$$

The maxima of intensity lie at :

$$\alpha = \frac{2\pi x}{\lambda r_0} = 0, \frac{2\pi}{c}, \frac{4\pi}{c}, \ldots \frac{2n\pi}{c}, \ldots \quad . \quad . \quad (230)$$

that is, at equal intervals which are directly proportional to the wave-length and inversely proportional to the grating constant. For axial light ($\theta = 0$) the maximum is common to all wave-lengths; a scattering of colours first shows itself at the next maximum, for $n = 1$, which gives the spectrum of the first order.

The value of the maximum is the same for all orders and is equal to $N^2 J_1$. We must take care to observe that the factor J_1 in (229), the diffracted intensity of a single aperture, is not constant but depends in its turn on the angle of diffraction and the width of slit b in the manner given by (219). Hence gratings with the same grating constants often differ considerably from one another in their effects. If, for example, the interval between two neighbouring slits is exactly as great as the slit width ($c = 2b$), the zero-points of J_1 in (220) coincide with the maxima of even order number n in (230), and hence all spectra of even order numbers are missing in the diffraction pattern of such a grating.

PART TWO

OPTICS OF CRYSTALS

CHAPTER I

PLANE WAVES

§ 52. THE optics of crystals is founded on the same electromagnetic field equations (1), (2) and (3) as the optics of isotropic bodies. The only difference is that the relationship between the electric induction $\boldsymbol{D}$ and the electric intensity of field $\boldsymbol{E}$, which is expressed in the case of isotropic substances by the simple proportionality (4), has a more general form in the case of crystals, in that the components of electric induction are certain linear homogeneous functions of the components of the electric intensity of field, that is :

$$\left.\begin{aligned} \boldsymbol{D}_x &= \epsilon_{11}\boldsymbol{E}_x + \epsilon_{12}\boldsymbol{E}_y + \epsilon_{13}\boldsymbol{E}_z \\ \boldsymbol{D}_y &= \epsilon_{21}\boldsymbol{E}_x + \epsilon_{22}\boldsymbol{E}_y + \epsilon_{23}\boldsymbol{E}_z \\ \boldsymbol{D}_z &= \epsilon_{31}\boldsymbol{E}_x + \epsilon_{32}\boldsymbol{E}_y + \epsilon_{33}\boldsymbol{E}_z \end{aligned}\right\} \quad . \quad . \quad (231)$$

where the constants ϵ depend on the nature of the crystal. The higher the degrees of symmetry the crystal has (II, § 26), the more relationships there are between these constants. For the limiting case of an isotropic substance $\epsilon_{11} = \epsilon_{22} = \epsilon_{33}$ and all the other ϵ's are equal to zero, because (231) then becomes transformed into (4).

But even when there is no symmetry at all only six of the nine quantities ϵ are independent of one another. This follows if we apply the energy principle to any electromagnetic process that occurs within the crystal. For if we fix our attention on any portion of space in the interior of the crystal and apply the energy principle to it, exactly as was done in III, § 9, we get for the amount of energy that flows through all the elements $d\sigma$ of the surface into space in the time dt, as in III (25) :

$$dt \,.\int d\sigma\, \mathbf{S}_\nu = -\,dt \int d\tau \,.\, \mathrm{div}\, \mathbf{S}$$

and, by applying (3) and (1), this equation is converted into :

$$\frac{dt}{4\pi} \cdot \int d\tau \, (\boldsymbol{E}_x\dot{\boldsymbol{D}}_x + \boldsymbol{E}_y\dot{\boldsymbol{D}}_y + \boldsymbol{E}_z\dot{\boldsymbol{D}}_z + \boldsymbol{H}_x\dot{\boldsymbol{H}}_x + \boldsymbol{H}_y\dot{\boldsymbol{H}}_y + \boldsymbol{H}_z\dot{\boldsymbol{H}}_z) \quad (232)$$

Hence by the energy principle this expression simultaneously also represents the change in the time dt of the electromagnetic energy contained in the part of the crystal under consideration. It consists, as we see, of two parts : electric and magnetic energy, as in the case of isotropic bodies. The time differential of the electric energy-density is :

$$\frac{1}{4\pi}(\boldsymbol{E}_x d\boldsymbol{D}_x + \boldsymbol{E}_y d\boldsymbol{D}_y + \boldsymbol{E}_z d\boldsymbol{D}_z). \quad . \quad . \quad (233)$$

that of the magnetic energy-density is :

$$\frac{1}{4\pi}(\boldsymbol{H}_x d\boldsymbol{H}_x + \boldsymbol{H}_y d\boldsymbol{H}_y + \boldsymbol{H}_z d\boldsymbol{H}_z) \quad . \quad . \quad (234)$$

Now whereas the magnetic energy-density, as a comparison with III, 3 shows, is exactly equal to that of an isotropic body of permeability $\mu = 1$, the electric energy-density has a considerably more involved form. For if we substitute (231) in (233) and then integrate we get a quadratic form in the components of the electric intensity of field, whose coefficients are composed of the constants ϵ. But if the integration is to be possible at all the coefficient of $\boldsymbol{E}_x d\boldsymbol{E}_y$ must be equal to that of $\boldsymbol{E}_y d\boldsymbol{E}_x$—that is, we must have :

$$\epsilon_{12} = \epsilon_{21}, \text{ and likewise } \epsilon_{23} = \epsilon_{32}, \ \epsilon_{31} = \epsilon_{13}.$$

Then the electric energy-density is :

$$\frac{1}{8\pi}(\epsilon_{11}\boldsymbol{E}_x^2 + \epsilon_{22}\boldsymbol{E}_y^2 + \epsilon_{33}\boldsymbol{E}_z^2 + 2\epsilon_{12}\boldsymbol{E}_x\boldsymbol{E}_y + 2\epsilon_{23}\boldsymbol{E}_y\boldsymbol{E}_z + 2\epsilon_{31}\boldsymbol{E}_z\boldsymbol{E}_x) \quad . \quad (235)$$

This expression may be considerably simplified by choosing as co-ordinate axes the three mutually perpendicular directions which form the axes of the ellipsoid represented by the tensor (235). These are the three

"principal axes" of the crystal. The electric energy then assumes the form:

$$\frac{1}{8\pi}(\epsilon_1 \boldsymbol{E}_x^2 + \epsilon_2 \boldsymbol{E}_y^2 + \epsilon_3 \boldsymbol{E}_z^2) \quad . \quad . \quad . \quad . \quad (236)$$

and the relation between the electric induction and electric intensity becomes more simple:

$$\boldsymbol{D}_x = \epsilon_1 \boldsymbol{E}_x, \ \boldsymbol{D}_y = \epsilon_2 \boldsymbol{E}_y, \ \boldsymbol{D}_z = \epsilon_3 \boldsymbol{E}_z \quad . \quad . \quad (237)$$

The three constants ϵ are called the "principal dielectric constants"; they are always positive.

In the sequel we shall use the principal axes as the co-ordinate axes throughout.

§ 53. We next propose the general question: are plane linearly polarized waves possible in a crystal? And if so, what laws do they obey? To answer this question we form the expressions for the field-components of such a wave and substitute them in the field-equations (1) and (2).

Let n be the normal to the plane wave, and α, β, γ its direction cosines, thus:

$$n = \alpha x + \beta y + \gamma z \quad . \quad . \quad . \quad . \quad . \quad (238)$$

Further, let D be the value of the electric induction, and ξ, η, ζ its direction cosines, so that:

$$\boldsymbol{D}_x = D \, . \, \xi, \ \boldsymbol{D}_y = D \, . \, \eta, \ \boldsymbol{D}_z = D \, . \, \zeta \quad . \quad . \quad (239)$$

and E the value of the electric intensity of field, and ξ', η', ζ' its direction cosines, so that:

$$\boldsymbol{E}_x = E \, . \, \xi', \ \boldsymbol{E}_y = E \, . \, \eta', \ \boldsymbol{E}_z = E \, . \, \zeta' \quad . \quad . \quad (240)$$

Lastly, let H be the value of the magnetic field-strength and λ, μ, ν its direction cosines, so that:

$$\boldsymbol{H}_x = H \, . \, \lambda, \ \boldsymbol{H}_y = H \, . \, \mu, \ \boldsymbol{H}_z = H \, . \, \nu \quad . \quad . \quad (241)$$

The condition for a plane linearly polarized wave is then that the quantities D, E and H, besides depending on the time t depend only on the length n, whereas the direction cosines that have been introduced are all constant.

With this assumption the field-equations (1) become transformed into the conditions :

$$\left.\begin{aligned} \xi\dot{D} &= c_0(\beta\nu - \gamma\mu)\cdot\frac{\partial H}{\partial n} \\ \eta\dot{D} &= c_0(\gamma\lambda - \alpha\nu)\cdot\frac{\partial H}{\partial n} \\ \zeta\dot{D} &= c_0(\alpha\mu - \beta\lambda)\cdot\frac{\partial H}{\partial n} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (242)$$

$$\left.\begin{aligned} \lambda\dot{H} &= c_0(\eta'\gamma - \zeta'\beta)\cdot\frac{\partial E}{\partial n} \\ \mu\dot{H} &= c_0(\zeta'\alpha - \xi'\gamma)\cdot\frac{\partial E}{\partial n} \\ \nu\dot{H} &= c_0(\xi'\beta - \eta'\alpha)\cdot\frac{\partial E}{\partial n} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (243)$$

which contain the complete answer to the questions above proposed. In them we have denoted the velocity of light *in vacuo* by c_0 as the symbol c will be used a little later (§ 55) with another significance.

§ 54. Concerning the directions of the various vectors which characterize a plane linearly polarized wave it follows from (242) that the electric induction is perpendicular both to the wave-normals and to the magnetic intensity of field :

$$\boldsymbol{D} \perp \boldsymbol{n} \quad \text{and} \quad \boldsymbol{D} \perp \boldsymbol{H} \quad . \quad . \quad . \quad (244)$$

Further, it follows from (243) that the magnetic intensity of field is perpendicular both to the wave-normals and the electric intensity of field :

$$\boldsymbol{H} \perp \boldsymbol{n} \quad \text{and} \quad \boldsymbol{H} \perp \boldsymbol{E} \quad . \quad . \quad . \quad (245)$$

These theorems also contain the results which are arrived at by substituting the assumed values for the field-components in the field-equations (2).

If we add to these the fact that, by (3), the flux-vector $\boldsymbol{S}$ is perpendicular to both the electric intensity of field and the magnetic intensity of field :

$$\boldsymbol{S} \perp \boldsymbol{E} \quad \text{and} \quad \boldsymbol{S} \perp \boldsymbol{H} \quad . \quad . \quad . \quad (246)$$

then all these relationships can be expressed in the single theorem that the magnetic intensity of field $\boldsymbol{H}$ is perpendicular to all the other vectors that have been mentioned; that is, the latter all lie in a plane, the electric induction $\boldsymbol{D}$ being perpendicular to the wave-normal $\boldsymbol{n}$ and the electric intensity of field $\boldsymbol{E}$ being normal to the ray $\boldsymbol{S}$. These relationships are represented graphically in Fig. 13, the plane of which has the magnetic intensity of field for its normal, its direction being towards the observer. This plane is called the "vibration plane," as in the case of isotropic bodies (§ 9). Thus the plane of vibration contains the normals to the wave, the ray, and directions of electric induction and electric intensity of field, which are perpendicular to the ray. The distinctive part played by the magnetic intensity of field in this relationship is, of course, due to the fact that it coincides with the magnetic induction, whereas this is not, in general, the case with the electric intensity and the electric induction.

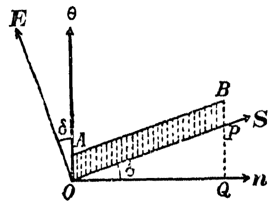

FIG 13.

If we denote the angle between the electric intensity of field and the electric induction δ, then δ is also the angle between the ray and the normal. Since the principal dielectric constants in (237) are positive, $\delta < \frac{\pi}{2}$.

The physical meaning of the difference between the directions of the ray and the normal may be made clear by the following considerations. If we imagine the assumed plane-wave to be laterally limited—which is effected by selecting any finite surface area OA (Fig. 13) in a wave-plane, this area represents the cross-section of a cylinder of light which propagates itself onwards from it. It is formed by rays which proceed from all points of the cross-section in an oblique direction, since they deviate from the direction of the normal by the angle δ. In

Fig. 13 the cross-section of the wave at different points of the cylinder, from OA to PB, is indicated by broken parallel lines.

The value of δ comes out from its definition as :

$$\cos\delta = \xi\xi' + \eta\eta' + \zeta\zeta' > 0 \quad . \quad . \quad . \quad (247)$$

or, if we also denote the direction cosines of the ray $\boldsymbol{S}$ by $\alpha'\beta'\gamma'$:

$$\cos\delta = \alpha\alpha' + \beta\beta' + \gamma\gamma' \quad . \quad . \quad . \quad (248)$$

Both the unaccented and the accented direction cosines form an orthogonal triplet with those of the magnetic intensity of field.

§ 55. We shall now pass from the directions to the actual magnitudes of the vectors in question. For this purpose we eliminate the direction cosines by squaring and adding the equations (242) and (243). The sum of the squares of the bracketed quantities in (242) is 1, since $\boldsymbol{H}\perp\boldsymbol{n}$. The same result is not, however, obtained for the corresponding sum in (243), since $\boldsymbol{E}$ is not $\perp\boldsymbol{n}$. Rather :

$$\begin{aligned}(\eta'\gamma - \zeta'\beta)^2 + (\zeta'\alpha - \xi'\gamma)^2 + (\xi'\beta - \eta'\alpha)^2 \\ = (\xi'^2 + \eta'^2 + \zeta'^2)\,.\,(\alpha^2 + \beta^2 + \gamma^2) - (\xi'\alpha + \eta'\beta + \zeta'\gamma)^2 \\ = 1 - \sin^2\delta = \cos^2\delta.\end{aligned}$$

Consequently we obtain the relationships :

$$\dot{D}^2 = c_0{}^2\left(\frac{\partial H}{\partial n}\right)^2 \quad \text{and} \quad \dot{H}^2 = c_0{}^2\left(\frac{\partial E}{\partial n}\right)^2\cdot\cos^2\delta \,. \quad (249)$$

In taking the square root, attention must be paid to the signs. They may be obtained from any particular case, for example, for that where the normal $\boldsymbol{n}$ lies in the x-axis, the electric induction $\boldsymbol{D}$ in the y-axis, and the magnetic intensity of field $\boldsymbol{H}$ in the z-axis. Then $\alpha = 1$, $\beta = 0$, $\gamma = 0$, $\xi = 0$, $\eta = 1$, $\zeta = 0$, $\lambda = 0$, $\mu = 0$, $\nu = 1$, $\eta' = \cos\delta$. Hence the equations (242) and (243) become :

$$\dot{D} = -\,c_0\frac{\partial H}{\partial n},\ \dot{H} = -\,c_0\,.\cos\delta\frac{\partial E}{\partial n} \quad . \quad . \quad (250)$$

and, by (249), these relationships also hold in the general case.

If we eliminate H by differentiating the first equation with respect to t and the second with respect to n, we get:

$$\frac{\partial^2 D}{\partial t^2} = c_0^2 \cos\delta \cdot \frac{\partial^2 E}{\partial n^2} \quad . \quad . \quad . \quad . \quad (251)$$

On the other hand, we get from (237).

$$D \,.\, \xi = \epsilon_1 \,.\, E\xi',\ D \,.\, \eta = \epsilon_2 \,.\, E\eta',\ D \,.\, \zeta = \epsilon_3 \,.\, E\zeta'.$$

If we take the value of D obtained from one of these three equations and substitute it in (251) we obtain the well-known wave-equation (6) in the form:

$$\frac{\partial^2 E}{\partial t^2} = q^2 \frac{\partial^2 E}{\partial n^2} \quad . \quad . \quad . \quad . \quad . \quad (252)$$

where q^2, the square of the velocity of propagation, has the following value:

$$q^2 = \frac{\xi}{\xi'} \frac{c_0^2}{\epsilon_1} \cos\delta = \frac{\eta}{\eta'} \frac{c_0^2}{\epsilon_2} \cos\delta = \frac{\zeta}{\zeta'} \frac{c_0^2}{\epsilon_3} \cos\delta$$

or, if we introduce the "principal velocities of propagation":

$$\frac{c_0^2}{\epsilon_1} = a^2,\ \frac{c_0^2}{\epsilon_2} = b^2,\ \frac{c_0^2}{\epsilon_3} = c^2 \quad . \quad . \quad . \quad (253)$$

$$\xi' q^2 = \xi a^2 \cos\delta,\ \eta' q^2 = \eta b^2 \cos\delta,\ \zeta' q^2 = \zeta c^2 \cos\delta \quad . \quad (254)$$

The differential equation (252) is satisfied, as we know from II, § 35, by the expression:

$$E = f\left(t - \frac{n}{q}\right) . \quad . \quad . \quad . \quad . \quad . \quad (255)$$

where f, the form of the wave, represents any arbitrary function of the single argument $t - \frac{n}{q}$. From this equation we further get:

$$\frac{\partial E}{\partial n} = -\frac{1}{q}\frac{\partial E}{\partial t} \quad . \quad . \quad . \quad . \quad . \quad (256)$$

and, by substituting in (250):

$$H = \frac{c_0}{q} \cdot \cos \delta \, . \, E \quad . \quad . \quad . \quad . \quad . \quad (257)$$

$$D = \frac{c_0^2}{q^2} \cdot \cos \delta \, . \, E \quad . \quad . \quad . \quad . \quad . \quad (258)$$

Hence, as soon as we have found a value for q which satisfies all three equations (254), we know that it is possible to have a wave of any form in the crystal with the corresponding direction of propagation and polarization.

§ 56. Let us now see whether it is possible to find expressions for the different constants which fulfil all the necessary conditions, above all, the equations (254).

In the first place, it is easy to see that if the direction of the electric induction $\boldsymbol{D}$ is assumed in any arbitrary way, all the other quantities result uniquely.

For if ξ, η ζ are known then by multiplying the individual equations (254) by ξ, η, ζ and adding, we get, in view of (247), that:

$$q^2 = a^2\xi^2 + b^2\eta^2 + c^2\zeta^2 \quad . \quad . \quad . \quad (259)$$

and, further, from (254), the direction of the intensity of field $\boldsymbol{E}$:

$$\xi' : \eta' : \zeta' = a^2\xi : b^2\eta : c^2\zeta \quad . \quad . \quad . \quad (260)$$

But the directions of $\boldsymbol{D}$ and $\boldsymbol{E}$ fix the vibration-plane and hence also the normal direction $\boldsymbol{H}$, as well as that of $\boldsymbol{n}$ and $\boldsymbol{S}$. So the wave is completely determined except for its form, which remains arbitrary.

In addition, we shall here derive only the expressions for the direction cosines α, β, γ of the wave-normals.

Since the wave-normal $\boldsymbol{n}$ lies in the same plane as the intensity of field $\boldsymbol{E}$ and the induction $\boldsymbol{D}$, the ratio of their direction cosines is:

$$\alpha : \beta : \gamma = (\xi' + x\xi) : (\eta' + x\eta) : (\zeta' + x\zeta)$$

where x denotes a certain number. But, on the other hand, the wave-normal is perpendicular to the induction. Hence:

$$\alpha\xi + \beta\eta + \gamma\zeta = 0 \quad . \quad . \quad . \quad . \quad . \quad (261)$$

whence we get, in view of (247):

$$x = -\cos\delta$$

Thus:

$$\alpha : \beta : \gamma = (\xi' - \xi\cos\delta) : (\eta' - \eta\cos\delta) : (\zeta' - \zeta\cos\delta)$$

from which we get, by (254):

$$\alpha : \beta : \gamma = (a^2 - q^2)\,.\,\xi : (b^2 - q^2)\,.\,\eta : (c^2 - q^2)\,.\,\zeta\,. \quad (262)$$

§ 57. Let us next inquire as to how far a plane wave is determined if the direction of the wave-normal $\boldsymbol{n}$ is given in any way. Our problem is then to express all quantities in terms of α, β, γ, and for this it is sufficient, as we have seen, to determine the direction cosines ξ, η, ζ. These come out by (262) as:

$$\xi : \eta : \zeta = \frac{\alpha}{a^2 - q^2} : \frac{\beta}{b^2 - q^2} : \frac{\gamma}{c^2 - q^2} \quad . \quad . \quad (263)$$

The velocity of propagation q, which still remains undetermined in this equation, is calculated according to (261) from the equation:

$$\frac{\alpha^2}{a^2 - q^2} + \frac{\beta^2}{b^2 - q^2} + \frac{\gamma^2}{c^2 - q^2} = 0 \quad . \quad . \quad (264)$$

As an exactly similar investigation in III, § 25, shows, this equation has two real positive roots in q^2, and hence also in q. They lie in the two intervals that are bounded by the values of the constants a, b, c. If we call them q_1 and q_2, we can, without restricting the generality of our case, agree that:

$$a \geqq q_1 \geqq b \geqq q_2 \geqq c \quad . \quad . \quad . \quad . \quad . \quad (265)$$

With each of the two values q_1 and q_2 there is associated, by (263), a definite direction of the electric induction $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$, and hence, by § 56, all the other directions are fixed.

The two directions of induction $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ that belong to a definite wave-normal α, β, γ, ν are very simply related to each other. For if in the equation (261) we substitute for ξ, η, ζ the values ξ_1, η_1, ζ_1, but for α, β, γ the values (262) with q_2 and ξ_2, η_2, ζ_2, we get:

$$(a^2 - q_2^2)\xi_1\xi_2 + (b^2 - q_2^2)\eta_1\eta_2 + (c^2 - q_2^2)\zeta_1\zeta_2 = 0.$$

Similarly :

$$(a^2 - q_1^2)\xi_1\xi_2 + (b^2 - q_1^2)\eta_1\eta_2 + (c^2 - q_1^2)\zeta_1\zeta_2 = 0.$$

From these we get by subtraction :

$$\xi_1\xi_2 + \eta_1\eta_2 + \zeta_1\zeta_2 = 0 \quad . \quad . \quad . \quad (266)$$

and :

$$a^2\xi_1\xi_2 + b^2\eta_1\eta_2 + c^2\zeta_1\zeta_2 = 0 \,. \quad . \quad . \quad (267)$$

The first equation states that $\boldsymbol{D}_1 \perp \boldsymbol{D}_2$, the second that $\boldsymbol{D}_1 \perp \boldsymbol{E}_2$ and that $\boldsymbol{D}_2 \perp \boldsymbol{E}_1$. From this we also obtain further that $\boldsymbol{H}_1 \parallel \boldsymbol{D}_1$ and $\boldsymbol{H}_2 \parallel \boldsymbol{D}_1$.

§ 58. The relationship between all these directions and the quantities of the corresponding velocities of propagation q_1 and q_2 are realized graphically by considering the so-called Cauchy "ellipsoid of polarization." This is the ellipsoid :

$$a^2x^2 + b^2y^2 + c^2z^2 = 1 \quad . \quad . \quad . \quad (268)$$

which has the semi-axes $\frac{1}{a}, \frac{1}{b}, \frac{1}{c}$.

If we wish to find the two waves that correspond to an arbitrary given normal direction $\boldsymbol{n}$, we draw the diametral plane of the ellipsoid perpendicular to $\boldsymbol{n}$. The directions of the two axes of the elliptic section then represent the directions of the electric inductions $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$, and the reciprocal lengths of the corresponding semi-axes represent the corresponding velocities of propagation q_1 and q_2.

For if instead of x, y, z we introduce a new rectangular co-ordinate system $x'y'z'$ having the same origin and with its z'-axis coinciding with the given normal direction $\boldsymbol{n}$ and its x'- and y'-axes coinciding with the directions of the electric inductions $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$, the equations of transformation run :

$$\begin{aligned} x &= \xi_1 x' + \xi_2 y' + \alpha z' \\ y &= \eta_1 x' + \eta_2 y' + \beta z' \\ z &= \zeta_1 x' + \zeta_2 y' + \gamma z' \end{aligned}$$

If we substitute these expressions in (268) and set

$z' = 0$ we get the equation of the elliptic section in view of (267) in the form .

$$(a^2\xi_1^2 + b^2\eta_1^2 + c^2\zeta_1^2)x'^2 + (a^2\xi_2^2 + b^2\eta_2^2 + c^2\zeta_2^2)y'^2 = 1$$

from which by (259) the values q_1 and q_2 result for the reciprocals of the semi-axes of the ellipse.

Since the magnetic intensity of field $\boldsymbol{H}$ is perpendicular to both $\boldsymbol{n}$ and $\boldsymbol{D}$ each of the two axes of the elliptic section simultaneously represents the direction of the magnetic intensity of field, which belongs to the direction of electric induction represented by the other axis, whereas the direction of the electric field-strength $\boldsymbol{E}$ is, by (260) given by the normal of the ellipsoid at the end-point of the axis which represents the electric induction. Finally the ray-direction $\boldsymbol{S}$ is represented by that tangent of the ellipsoid which lies in the vibration-plane defined by $\boldsymbol{n}$, $\boldsymbol{D}$ and $\boldsymbol{E}$.

If a principal axis of the crystal, say the x-axis, is a normal to the wave, then the other two axes y and z are the corresponding directions of the induction with the velocities of propagation b and c.

All these theorems, of course, remain valid for the case where $a = b = c$. The body is then isotropic optically and the polarization ellipsoid is a sphere, the elliptic section is a circle, every diameter is an axis and the velocity of propagation $q = a$.

§ 59. Whereas hitherto we have assumed α, β, γ as given and q_1 and q_2 as determined by them, the case may also arise where q_1 and q_2 are prescribed initially, of course within the limits (265), and the corresponding values α, β, γ of the wave-normals are to be found.

To solve this problem it is useful to write the equation (264) in the form :

$$\alpha^2(b^2 - q^2)(c^2 - q^2) + \beta^2(c^2 - q^2)(a^2 - q^2) + \gamma^2(a^2 - q^2)(b^2 - q^2)$$
$$= (q^2 - q_1^2) \cdot (q^2 - q_2^2) \quad . \quad . \quad . \quad . \quad (269)$$

This equation is an identity in so far as it holds, for any arbitrary value of q, provided only that q_1 and q_2 are the roots of (264).

According as we now set $q = a$, b or c, we get from (269):

$$\left.\begin{aligned} \alpha^2 &= \frac{(a^2 - q_1^2) . (a^2 - q_2^2)}{(b^2 - a^2) . (c^2 - a^2)} \\ \beta^2 &= \frac{(b^2 - q_1^2) . (b^2 - q_2^2)}{(c^2 - b^2) . (a^2 - b^2)} \\ \gamma^2 &= \frac{(c^2 - q_1^2) . (c^2 - q_2^2)}{(a^2 - c^2) . (b^2 - c^2)} \end{aligned}\right\} \quad . \quad . \quad (270)$$

These relationships enable α^2, β^2, γ^2 to be determined as soon as q_1 and q_2 are given. It is easy to convince oneself that by (265) the direction cosines α, β, γ are always real. They denote, according to their signs, eight different directions, each of which lies in a particular octant of the co-ordinate system of the principal axes, or, if we take any two in opposite directions as standing for one, we get four different directions.

Then, by § 58, the values of α, β, γ also denote the direction of the electric induction and so forth that belong to q_1 and to q_2.

The special case where $q_1 = q_2$ is of particular interest. For by (265) we then have:

$$q_1 = q_2 = b \quad . \quad . \quad . \quad . \quad . \quad (271)$$

and, by (270), the direction cosines of the corresponding normals come out as:

$$\alpha_0^2 = \frac{a^2 - b^2}{a^2 - c^2}, \ \beta_0 = 0, \ \gamma_0^2 = \frac{b^2 - c^2}{a^2 - c^2} \quad . \quad . \quad (272)$$

Of these four directions that lie in the xz-plane those two which form an acute angle with the z-axis—namely, the axis of the smallest principal velocity of propagation, for which the following values hold:

$$\alpha_0 = \pm \sqrt{\frac{a^2 - b^2}{a^2 - c^2}}, \ \beta_0 = 0, \ \gamma_0 = \sqrt{\frac{b^2 - c^2}{a^2 - c^2}} \quad . \quad (273)$$

are called the "optic axes" of the crystal. The crystal is called "optically positive" if the two optic axes form

an acute angle with each other—that is, when $\gamma_0 > \sqrt{\frac{1}{2}}$ or

$$b^2 > \frac{a^2 + c^2}{2} \quad . \quad . \quad . \quad . \quad . \quad (274)$$

In the contrary case, the crystal is called "optically negative."

In Cauchy's ellipsoid of polarization the optic axes denote the normals of the two diametral planes which cut the ellipsoid in a circle of radius $\frac{1}{b}$. Since the circle has an infinite number of axial directions, a wave whose normals coincide with that of the optic axis has an infinite number of directions for the electric induction, just as in the case of an isotropic body. Consequently we may say that with respect to an optic axis as wave-normal a crystal behaves just like an isotropic body. But an essential difference is that in an isotropic body the electric intensity of field coincides with the electric induction and consequently also the ray with the wave-normal in this direction, whereas in the case of a crystal the electric intensity of field that is, the normal of the ellipsoid of polarization at a point of a circular intersection in general by no means coincides with the electric induction that is, with the radius of the circle at this point. Rather, each of the infinite number of directions of induction has its own particular field-strength and its own particular ray direction perpendicular to its field-strength, so that for each definite optic axis regarded as a wave-normal there are an infinite number of different rays, constituting a whole cone of rays. Among the generators of this cone there is also to be found the optic axis itself. For at the point where the circle of intersection meets the y-axis the normal of the ellipsoid coincides with the radius of the circle, and hence the electric intensity of field with the electric induction, and consequently also the ray with the wave-normal.

If two of the three principal velocities of propagation a, b, c are equal the crystal is called uniaxial, being

"positive," by (274), if $a = b > c$, and "negative" if $a > b = c$.

In numerical calculations it is customary to use, not the values of the principal velocities of propagation, but those of the "principal refractive indices." Referred to a vacuum these are:

$$\frac{c_0}{a} = n_1, \ \frac{c_0}{b} = n_2, \ \frac{c_0}{c} = n_3 \quad . \quad . \quad . \quad (275)$$

or, by (253):

$$n_1 = \sqrt{\epsilon_1} \qquad n_2 = \sqrt{\epsilon_2} \qquad n_3 = \sqrt{\epsilon_3} \quad . \quad . \quad (276)$$

where, by (265):

$$n_1 \leqq n_2 \leqq n_3$$

Thus for $n_1 = n_2$ the crystal is positively uniaxial, and for $n_2 = n_3$ it is negatively uniaxial.

The optical properties of a crystal are closely related to its elastic structure; for every crystallographic symmetry also entails an optical symmetry. Hence the crystals of the non-symmetrical, monosymmetrical and rhombic system (II, § 26) are optically biaxial (topaz positive; aragonite and mica negative); those of the hexagonal and tetragonal systems are optically uniaxial (quartz, ice, zirconium positive, calcspar, turmaline negative); those of the regular system are optically isotropic (rock-salt, sylvine).

§ 60. The physical significance of the velocity of propagation q is best seen graphically in Fig. 13 (§ 54), in which the wave-plane OA is depicted as having moved to PB in unit time. The magnitude of q is then represented by the corresponding increase in the normal n; that is, by the distance OQ. If instead of measuring the velocity of propagation along the normal direction $\mathbf{n}$ we measure it along the ray-direction $\mathbf{S}$, we get the distance OP, which we appropriately designate by q'.

The general relation:

$$\frac{q}{q'} = \frac{OQ}{OP} = \cos \delta \quad . \quad . \quad . \quad . \quad . \quad (277)$$

then holds.

The appropriateness of our nomenclature is shown in the circumstance that each of the different relationships that have been set up between the direction cosines introduced and the velocities of propagation retain their validity if we exchange the accented and the unaccented direction cosines with each other and in addition replace q by $\frac{1}{q'}$, q' by $\frac{1}{q}$, and a, b, c by $\frac{1}{a}, \frac{1}{b}, \frac{1}{c}$. The direction of the magnetic intensity of field λ, μ, ν and the angle δ then remain unchanged in the transformation.

We have immediate proof of the correctness of this theorem if we consider that by § 54 the same laws hold for the accented direction cosines as for the unaccented, and that the equations (254) to (277), which serve to determine the velocity of propagation q, remain correct if we execute the specified exchanges in them.

In this behaviour we see a dualism, a general law of reciprocity, which is valid throughout and which may often be profitably used to derive new relationships between the properties of a wave. With its help we can immediately enunciate a number of theorems which are no less important than those which have preceded. We shall here mention a few of the most important of them.

Corresponding to every ray-direction α', β', γ' there are two velocities of propagation q'_1 and q'_2, where:

$$a \gtrless q'_1 \gtrless b \gtrless q'_2 \gtrless c \quad . \quad . \quad . \quad . \quad (278)$$

They are the two roots of the equation:

$$\frac{a^2\alpha'^2}{a^2 - q'^2} + \frac{b^2\beta'^2}{b^2 - q'^2} + \frac{c^2\gamma'^2}{c^2 - q'^2} = 0 \quad . \quad . \quad (279)$$

These lengths represent the semi-axes of the ellipse which is the intersection of the diametral plane perpendicular to α', β', γ' with the ellipsoid:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \quad . \quad . \quad . \quad . \quad . \quad (280)$$

The directions of the two axes give the directions of the corresponding electric intensities of field, whereas the

direction of the electric induction is indicated by the normal of the ellipsoid at the end of an axis, and the direction of the wave-normal by the tangent at the point of the ellipsoid in the vibration-plane defined by the ray and the electric intensity of field. The magnetic intensity of field forms the normal of the vibration plane, and so coincides with the other axis of the ellipse.

Conversely we have for any two arbitrarily given values q'_1 and q'_2, which fulfil the conditions (278), eight ray directions which are oppositely directed in pairs. They are represented by the values :

$$\left.\begin{aligned} \alpha'^2 &= \frac{b^2c^2}{q'^2_1 q'^2_2} \cdot \frac{(a^2 - q'^2_1) \cdot (a^2 - q'^2_2)}{(a^2 - b^2) \cdot (a^2 - c^2)} \\ \beta'^2 &= \frac{c^2a^2}{q'^2_1 q'^2_2} \cdot \frac{(b^2 - q'^2_1) \cdot (b^2 - q'^2_2)}{(b^2 - c^2) \cdot (b^2 - a^2)} \\ \gamma'^2 &= \frac{a^2b^2}{q'^2_1 q'^2_2} \cdot \frac{(c^2 - q'^2_1) \cdot (c^2 - q'^2_2)}{(c^2 - a^2) \cdot (c^2 - b^2)} \end{aligned}\right\} \quad . \quad . \quad (281)$$

If q'_1 and q'_2 can be set equal to each other, that is, if by (278) :

$$q'_1 = q'_2 = b \quad . \quad . \quad . \quad . \quad . \quad (282)$$

we get from (281) the two singular ray-directions which are called the " secondary optic axes " of the crystal :

$$\left.\begin{aligned} \alpha'_0 &= \pm \frac{c}{b} \cdot \sqrt{\frac{a^2 - b^2}{a^2 - c^2}} = \frac{c}{b} \cdot \alpha_0 \\ \beta'_0 &= 0 = \beta_0 \\ \gamma'_0 &= \frac{a}{b} \cdot \sqrt{\frac{b^2 - c^2}{a^2 - c^2}} = \frac{a}{b} \gamma_0 \end{aligned}\right\} \quad . \quad . \quad (283)$$

These are the normals of the two diametral planes that pass through the y-axis and which cut the ellipsoid (280) in circles of radius b. Comparison with the direction cosines α_0, β_0, γ_0 of the primary optic axes shows that their angle with the z-axis is smaller—that is, it is enclosed by primary optic axes. In uniaxial crystals they coincide with the primary axes, for positive crystals ($a = b$) with the z-axis, for negative crystals ($b = c$) with

the positive and negative x-axis. To a secondary optic axis as a ray there belong a whole cone of wave-normals, the generators of which include the ray itself. Corresponding to each of these normals there is a definite vibration-plane which passes through it and the ray and which contains the electric intensity of field and the electric induction.

All these theorems can, of course, also be deduced directly without the use of the law of reciprocity if we repeat the line of reasoning used from § 56 onwards by considering the unaccented quantities to be everywhere replaced by the accented quantities.

CHAPTER II

WAVE SURFACE

§ 61. The realisation of a plane wave is impossible in nature if only on account of its unlimited cross-section. But by applying the same considerations that were used in dealing with isotropic bodies (§ 5) we are able to realize plane waves to a degree of approximation that can be carried as far as we please. We take a point-source of light somewhere in the interior of a crystal which we suppose of any sufficiently great extent and assume that it begins to emit light at the moment $t = 0$ and we call the surface to which the light has advanced after the time t (which entirely surrounds O) the wave-surface belonging to the point O as centre. If we choose t sufficiently large we can mark off on the wave-surface a portion of any magnitude whose dimensions will yet be small compared with its distance from O and which may therefore be regarded as nearly plane. The whole wave-surface is then composed of facets, as it were, of these plane portions, for each of which the laws of propagation of plane-waves hold.

If there is difficulty in imagining the point-source of light in the interior of the crystal we may suppose the source to be situated at a point in the surface of the crystal, from which it propagates light into the interior. But in this case we do not obtain the completely closed wave-surfaces, but only a part cut out by the surface of the crystal; this does not, however, affect our subsequent arguments, as they refer essentially only to the consideration of a single plane portion of the wave.

Such a plane portion of a wave propagates itself in the

crystal, in accordance with the law which we deduced in the preceding chapter, with a velocity and a polarization which correspond with its orientation with respect to the principal axes of the crystal, exactly like the limited transverse section of an absolutely plane wave such as we indicated by PB in Fig. 13 (§ 54). In this case OP or OB, since $PB << OP$, denotes the direction of the ray and OQ the direction of the wave normal; and we have $OP = q' . t$, $OQ = q . t$. The fact that in the neighbourhood of the light source O the wave surface can no longer be supposed to consist of sufficiently great plane portions causes no difficulty since we can always choose the time t so great that the distances OP and OQ can be set proportional to the time t.

The above reflections furnish two different methods of determining completely the wave-surface corresponding to a sufficiently great time, say $t = 1$.

The first method uses the ray velocity. Let us draw the distance q' in any arbitrary direction $\alpha'\beta'\gamma'$. The end-points P of all these distances form the required wave-surface. The second method uses the normal velocity. We draw the distance q in any arbitrary direction $\alpha\beta\gamma$ and construct the plane perpendicular to this distance through its end-point Q. Then all the planes that are constructed in this way form the tangential planes of the required wave-surface. The first method gives us the equation of the wave-surface in point-co-ordinates, the second method gives us this equation in plane-co-ordinates.

It is possible to establish some properties of the wave-surface on the basis of these constructions, without having to make particular calculations. Since for every ray there are two values of the velocity of propagation, namely q'_1 and q'_2, every straight line drawn from O as starting-point cuts the wave-surface in two points P_1 and P_2; that is, the wave-surface consists of two shells, the outer of which is composed of the points P_1 and the inner of the points P_2. The tangential planes of the wave-surface at P_1 and P_2 are the two wave-planes that belong to the ray OP_1P_2,

their normals being the wave-normals. But there are four singular directions, namely the secondary optical axes (283) together with their opposite directions, which give only a single point of intersection with the wave-surface, at the distance b from O. Thus the wave-surface has four singular points situated in the xz-plane, at which the two shells meet. Since corresponding to each of the singular rays there are an infinite number of normal directions, the wave surface has an infinite number of tangential planes at each singular point n, that is, a whole cone of tangents, which give the surface in the immediate neighbourhood of the point a funnel-like appearance. This causes the outer shell to appear dented inwards at the point in question, and the inner shell to appear to have an outward bulge. The secondary optic axis itself is also a generator of the tangential cone.

On the other hand, corresponding to any arbitrary direction of the wave-normals there are two tangential planes of the wave-surface at the distances q_1 and q_2 from O; these planes are perpendicular to the given direction. The lines connecting the point of contact with O represent the directions and the velocities of propagation of the two associated rays. But there are four singular tangential planes, perpendicular to the optic axes (273) and at the distance b from O, for which the distances q_1 and q_2 both coincide with b. Corresponding to each of such singular tangential planes there are an infinite number of rays, and hence also an infinite number of points of contact with the wave-surface; that is, the plane touches the wave-surface in a curve. Hence the above-mentioned funnel has a plane edge; it can be closed completely by a plane sheet. The optic axis itself is also a member of the generators of the cone of rays marked out by the curve of contact.

All the above theorems, which follow of necessity from the developments of the last section, will be confirmed and illustrated graphically in the following special applications.

§ 62. We shall now build up the equation of the optical wave-surface in the point-co-ordinates x, y, z for $t = 1$; for this purpose we shall use the first of the two above-mentioned methods. According to this a point of the wave-surface is represented by the co-ordinates :

$$x = r\alpha', \ y = r\beta', \ z = r\gamma'$$

where :

$$r = \sqrt{x^2 + y^2 + z^2} = q'.$$

If we substitute these values of α', β', γ', q' in equation (279) we obtain as the equation of the optical wave-surface :

$$\frac{a^2x^2}{a^2 - r^2} + \frac{b^2y^2}{b^2 - r^2} + \frac{c^2z^2}{c^2 - r^2} = 0 \quad . \quad . \quad (284)$$

The surface has O, of course, as centre and divides into eight symmetrical octants. To obtain its order we multiply (284) by the common denominator. We then get an expression of the sixth order, but easily see that it contains the factor r^2. If we omit this factor, we get from (284) :

$$a^2b^2c^2 - \{a^2(b^2 + c^2)x^2 + b^2(c^2 + a^2)y^2 + c^2(a^2 + b^2)z^2\} + (x^2 + y^2 + z^2)(a^2x^2 + b^2y^2 + c^2z^2) = 0 \quad . \quad . \quad (285)$$

The surface is therefore of the fourth order and this agrees with the fact that a ray which starts from O cuts the surface in four points.

The form of the optical wave-surface is best shown graphically by considering its sections with the three principal planes. Let us first take the xy-plane. For $z = 0$ (285) becomes :

$$(a^2b^2 - a^2x^2 - b^2y^2)(c^2 - x^2 - y^2) = 0 \quad . \quad . \quad (286)$$

The curve of intersection of the wave-surface with the xy-plane thus resolves into the ellipse :

$$\frac{x^2}{b^2} + \frac{y^2}{a^2} = 1$$

whose semi-axes are b and a, and into the inscribed circle :

$$x^2 + y^2 = c^2$$

of radius c. It is represented in Fig. 14. The points (P_1) of the ellipse form the trace of the outer shell, the points (P_2) of the circle form the trace of the inner shell. The wave-surfaces corresponding to a ray **S** which has the points of intersection P_1 and P_2 are the tangential planes of the surface at P_1 and P_2, which are, of course, perpendicular to the xy-plane. This simultaneously determines the two mutually perpendicular planes of vibration which belong to the ray **S** and which pass through it and the wave-normal n. The vibration-plane at P_1 is the xy-plane, that at P_2, where the ray and the normal

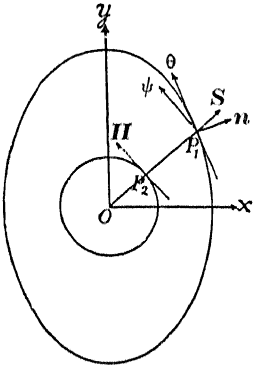

FIG. 14.

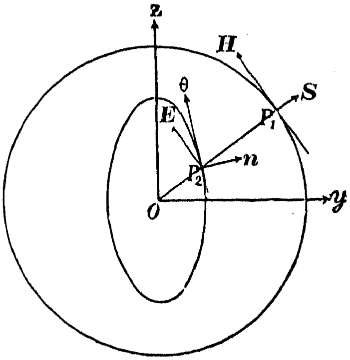

FIG. 15.

coincide, is the perpendicular plane. The vibration-plane contains the electric intensity of field **E** and the electric induction **D**, the first being perpendicular to **S** and the second to **n**. At P_1 they both lie in the plane of the diagram (see figure). At P_2 they are both perpendicular to this plane and coincide. Conversely, the magnetic intensity of field **H** at P_1 is perpendicular to the diagram, whereas at P_2 it lies in the plane of the diagram.

Exactly corresponding remarks may be made about the curves of intersection of the wave-surface with the other two principal planes. Let us first consider the zy-plane. For $x = 0$ (285) becomes:

$$(b^2c^2 - b^2y^2 - c^2z^2)\ (a^2 - y^2 - z^2) = 0 \quad . \quad . \quad (287)$$

We therefore obtain as the trace of the external shell the circle :

$$y^2 + z^2 = a^2$$

of radius a, and as the trace of the inner shell the ellipse :

$$\frac{y^2}{c^2} + \frac{z^2}{b^2} = 1$$

whose semi-axes are c and b. The corresponding diagram is shown in Fig. 15. It differs from that shown in Fig. 14 only in this respect, that for the points P_2 of the inner shell the vibration-plane falls in the plane of the diagram, whereas for the points P_1 of the outer shell it is perpendicular to that plane.

§ 63. The most interesting cross-section of the wave-surface is that made by the xz-plane, that with the greatest and least principal velocity of propagation. For the two spherical sections which the plane $y = 0$ cuts out of the wave-surface (285) :

$$(c^2a^2 - c^2z^2 - a^2x^2)(b^2 - z^2 - x^2) = 0 \quad (288)$$

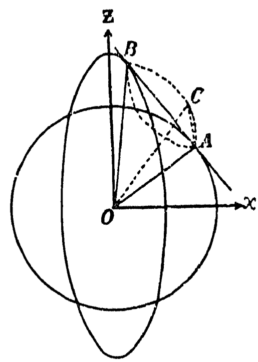

FIG. 16.

that is, the ellipse with the semi-axes c and a, and the circle of radius b, intersects in four real points, as is shown in Fig. 16. Since for every ray drawn from O to any such point the two velocities of propagation q'_1 and q'_2 corresponding to it coincide with each other and with b, those two of these four directions which make an acute angle with the z-axis form the two secondary optic axes of the crystal, and the trace of the outer shell of the wave-surface is partly represented by the two elliptic arcs which project beyond the circle and partly by the two circular arcs that project beyond the ellipse. The converse holds for the inner shell.

Like the secondary optic axes so the primary optic

axes also lie in the xz-plane, and, in fact, they form the normals of those tangential planes which touch the wave-surface along a whole curve. Their distance from O amounts to $q_1 = q_2 = b$, the angle between the primary optic axis is, in accordance with (283), greater than that between the secondary optic axes. The curve of contact represents the plane edge of the funnel, described at the end of § 61, which intersects the plane of the diagram at the points A and B and for the rest runs perpendicularly to this plane; for this reason it is depicted only by a dotted line in Fig. 16. The straight lines drawn from O to the points of the curve of contact constitute the cone of rays that belong to the optical axis OA; the generators of this cone also include the optic axis.

To find the form of the curve of contact we imagine the point C on the curve in the figure to be movable and starting from A to pass along the front arc of the curve to B and then to travel along the back arc of the curve to A again. In any position of C the corresponding ray is denoted by OC and hence the associated vibration-plane is denoted by the plane OAC and the corresponding electric induction by the direction AC which is perpendicular to OA and lies in the plane OAC.

The length of the distance AC is obtained from the right-angled triangle OAC and comes out as:

$$\begin{aligned} AC &= OA \,.\, \tan.\, AOC \\ &= b \,.\, \tan\delta = b \,.\, \frac{\sin\delta}{\cos\delta} = b\, \frac{\cos\left(\frac{\pi}{2} - \delta\right)}{\cos\delta} \end{aligned}$$

or, by (247) and since the wave-normal α_0, β_0, γ_0 is perpendicular to the electric induction ξ, η, ζ:

$$AC = b \cdot \frac{|\,\alpha_0\xi' + \beta_0\eta' + \gamma_0\zeta'\,|}{\xi\xi' + \eta\eta' + \zeta\zeta'}$$

where:

$$\alpha_0\xi + \beta_0\eta + \gamma_0\zeta = 0 \quad . \quad . \quad . \quad . \qquad (289)$$

From this it follows by (260) that:

$$AC = b \cdot \frac{|a^2\alpha_0\xi + b^2\beta_0\eta + c^2\gamma_0\zeta|}{a^2\xi^2 + b^2\eta^2 + c^2\zeta^2}$$

and, by (259) and (289), we get, taking into account that here $q = b$ and $\beta_0 = 0$:

$$AC = \frac{a^2 - c^2}{b}\gamma_0\zeta \quad . \quad . \quad . \quad . \quad (290)$$

The required curve of contact is completely determined by this relationship. For from it we get for every value of the direction cosine ζ and hence also for every direction of the straight line that rotates about A the corresponding value for the distance AC. If C coincides with A then $AC = 0$ and $\zeta = 0$; that is, the tangent of the curve at A is perpendicular to the z-axis, and since it is also perpendicular to the optic axis OA it lies in the direction of the y-axis, perpendicular to the plane of the diagram. If C then moves towards B the angle between AC and the z-axis decreases and so ζ increases up to a maximum which is attained when C coincides with B. For then, since AB is perpendicular to OA, we have $\zeta = \alpha_0$, and by (290):

$$AB = \frac{a^2 - c^2}{b}\alpha_0\gamma_0 \quad . \quad . \quad . \quad . \quad (291)$$

When the point C returns along the other arc of the curve the same values are traversed in the reverse direction until C again coincides with A. The relation (290) becomes still simpler if we introduce the angle CAB in place of ζ. This is the angle between the direction AC, which has the direction cosines ξ, η ζ and the direction AB, which has the direction cosines $-\gamma_0$, 0, α_1. So:

$$\cos CAB = -\gamma_0\xi + \alpha_0\zeta$$

and by (289):

$$\cos CAB = \left(\frac{\gamma_0^2}{\alpha_0} + \alpha_0\right)\zeta = \frac{\zeta}{\alpha_0}$$

Combined with (290) and (291) this gives:

$$AC = AB \, . \cos CAB$$

Hence it follows that the triangle ABC is right angled at C, and that the curve of contact ACB is a *circle*. But this does not imply that the cone of rays OC is a circular cone. For the circular cross-section ACB is perpendicular to a generator OA of the cone. We get as the simplest expression for the size of the aperture of the cone of rays:

$$\tan \omega = \frac{AB}{AO} = \frac{a^2 - c^2}{b^2} \alpha_0 \gamma_0$$

where ω denotes the angle between the rays OA and OB. And by (273) we get:

$$\tan \omega \frac{\sqrt{(a^2 - b^2)(b^2 - c^2)}}{b^2} \quad . \quad . \quad . \quad (292)$$

Exactly corresponding theorems may be derived from the law of reciprocity (§ 60) for the secondary optic axes and for the singular points of the optic wave-surface. At every singular point there are, corresponding to the singular ray connecting O to it, an infinite number of wave-planes and wave-normals, which form a cone of aperture ω', where, analogously to (292):

$$\tan \omega' = b^2 . \sqrt{\left(\frac{1}{a^2} - \frac{1}{b^2}\right)\left(\frac{1}{b^2} - \frac{1}{c^2}\right)}$$

or:

$$\tan \omega' = \frac{\sqrt{(a^2 - b^2)(b^2 - c^2)}}{ac} \quad . \quad . \quad . \quad (293)$$

CHAPTER III

NORMAL INCIDENCE

§ 64. SINCE the reflection and the refraction of light at the surface of a crystal obey considerably more complicated laws than in the case of isotropic bodies it seems advisable to begin here by considering some rather simple cases. We shall first take the simplest case, where a monochromatic plane light-wave in air falls normally on the surface of a crystal which we shall suppose to have any arbitrary direction with respect to its principal axes.

Since according to our assumption all points of the surface of the crystal are excited simultaneously in the same phase by the incident wave, the wave-plane and the wave-normal remain preserved also in the interior of the crystal. A peculiar feature, however, presents itself in that within the crystal there are two waves that propagate themselves with quite different velocities of propagation q_1 and q_2, and they also have quite different directions for their field-strengths and inductions as obtained from the laws derived above for a definite direction of the wave-normals. On emerging from the crystal each of the two waves obeys its own laws of refraction and reflection. Hence we may use the difference in the two velocities of propagation in the crystal to separate the waves entirely from one another, for example, by allowing one of them to be totally reflected when it impinges on the opposite surface of the crystal, while the other emerges refracted in the usual way. This is effected in a Nicol prism, which furnishes the simplest method of obtaining linearly polarized light from natural light.

For our discussion we shall now take a polished plane parallel crystal plate, so that even when the wave-planes emerge from the crystal they remain parallel to the surface. Since the electric inductions of the two waves are perpendicular to each other and to the wave-normals, they lie in the surface of the crystal and are also called the principal sections (optical) of the plate. They represent the wave-planes of the two waves and are denoted by *I* and *II* in the plane diagram of Fig. 17, which is to be imagined as parallel to the face of the plate. We shall suppose the light-waves to enter the plate from the rear and to pass out in the direction pointing towards the observer.

If the incident light is natural light the transmitted light will also be natural light if, as we shall do here, we disregard losses due to reflection. But the result is different if we allow polarized light to fall on the plate, say by allowing the light first to pass through a Nicol prism, which is in this case called a *polarizer*. The vibration-plane of the light which has been polarized in this way and is incident on the crystal is denoted in Fig. 17 by OP; let it form an angle ϕ with the principal section I. On entering the crystal the light-wave resolves into two waves whose vibration-planes are I and II, whose velocities of propagation are q_1 and q_2, whose phase-difference is zero, and whose amplitudes have the ratio $\cos \phi : \sin \phi$ (§ 22). When these two waves have passed through the crystalline plate, of thickness D, they have the following difference of phase on emerging from the plate:

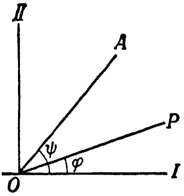

FIG. 17.

$$\Delta = \omega\left(\frac{D}{q_1} - \frac{D}{q_2}\right) = \frac{2\pi}{\lambda_0}(n_1 - n_2)D \quad . \quad . \quad (294)$$

where ω denotes the frequency, λ_0 the wave-length in air, n_1 and n_2 the indices of refraction. Hence in the air they in general combine to form an elliptically polarized

wave (§ 26), the directions of whose axes depend essentially on the quantity Δ. It is only when Δ is a whole multiple of π that linearly polarized light again results. If Δ is an odd multiple of $\frac{\pi}{2}$, the axes of the ellipse fall in the directions of the principal sections I and II; if, in addition, $\phi = \frac{\pi}{4}$ or $\frac{3\pi}{4}$ the light is circularly polarized. Hence a crystalline plate of this kind and of suitable thickness (for example, a thin sheet of mica) provides us with a simple means of transforming linearly polarized light into circularly polarized light, or vice versa by being used as a compensator, and of analysing elliptically polarized light into its rectilinear components.

If we make the light which emerges from the crystal plate in the forward direction pass through a second Nicol prism acting as an *analyser*, its direction of vibration OA (Fig. 17) making an angle ψ with the principal section I, then of the wave I with amplitude $\cos\phi$ only the component $\cos\phi \,.\, \cos\psi$ and of the wave II with amplitude $\sin\phi$ only the component $\sin\phi \,.\, \sin\psi$ pass through the analyser; and these two waves, both of which vibrate in the direction OA, give the following intensity, according to (101), the light delivered by the polarizer being taken to be of unit intensity:

$$J = \cos^2\phi\cos^2\psi + \sin^2\phi\sin^2\psi + 2\cos\phi\cos\psi\sin\phi\sin\psi\cos\Delta$$

or:

$$J = \cos^2(\phi - \psi) - \sin 2\phi \sin 2\psi \sin^2\frac{\Delta}{2} \quad . \quad . \quad (295)$$

A particularly interesting case is that where the directions of vibration OA and OP are mutually perpendicular, that is, where the polarizer and the analyser are in the "crossed" position. For then $\psi = \phi + \frac{\pi}{2}$, and from (295) we get:

$$J = \sin^2 2\phi \,.\, \sin^2\frac{\Delta}{2} \quad . \quad . \quad . \quad . \quad (296)$$

In this case J vanishes simultaneously with Δ. Hence this proves that whenever light can pass through a plate placed between two crossed Nicols the phase-difference Δ and hence also $q_1 - q_2$ differ from zero. This gives us one of the simplest means of testing whether a substance is anisotropic or not. The directions of the principal sections I and II can also be directly determined then, since by (296) the transmitted light has its maximum intensity for $\phi = \frac{\pi}{4}$ and $\phi = \frac{3\pi}{4}$, whereas for $\phi = 0$ and $\phi = \frac{\pi}{2}$ it vanishes entirely, as is quite natural.

§ 65. So far in dealing with light-waves in the crystal we have considered only the wave-normals and not the rays. This is allowable so long as the cross-section of the wave may be assumed to be unlimited. But so soon as the cross-section of the wave is limited the circumstance asserts itself that the energy of the wave does not propagate itself in the direction of the wave-normal but in that of the ray as has been explained earlier, in § 54, by means of Fig. 13.

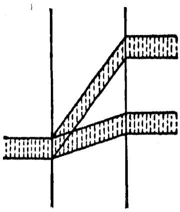

FIG. 18.

Now since in general there are two different rays for a definite wave-normal the two waves into which the plane beam of rays that falls normally into the plate resolves itself will sooner or later, depending on the size of its cross-section, separate and emerge from the opposite side of the crystal as two distinct beams of rays, as is indicated in Fig. 18. In this sense the crystal plate is doubly refracting even when the light is normally incident, although the wave-planes in all the rays, inside and outside the crystal, always retain the same direction. Here then the double-refraction is not, as in the case of isotropic bodies, determined by the difference between the substances in contact but by the nature of the crystal alone.

§ 66. A particularly interesting case occurs when the crystal plate, say of arragonite, is cut perpendicularly to the optic axis. For then an infinite number of rays

belong to the one wave-plane parallel to the surface of the plate, and the light, on entering the crystal, propagates itself not in two, but, if it was originally unpolarized, in all the directions of the generators of the ray-cone which corresponds to the optic axis as normal. The optic axis is also included among them. The plane of the diagram of Fig. 19, like that of Fig. 16 in § 36, contains besides the optic axis also the axis of the greatest and the smallest principal velocity of propagation (xz-plane). The directions OA and OB denote, as in Fig. 16, the optic-axis and the ray of the cone that diverges most from it. Since the section of the surface of the plate, that is, the singular wave-plane with the ray-cone, is a circle, the rays that pass out into the air form the mantle of a circular cylinder, whose diameter AB is obtained, according to Fig. 19, from the angle of aperture ω of the ray-cone and the thickness $OA \pm D$; its value is $D \tan \omega$, where the value (292) must be inserted for $\tan \omega$. The phenomenon which has just been described is one of the most brilliant confirmations of theoretical crystalline optics; it is called "conical refraction," and in the particular case considered "internal conical refraction," because the conical resolution of the light occurs when the light enters into the crystal.

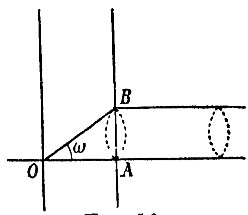

FIG. 19.

§ 67. As in the case of every law that refers to the ray-direction in the crystal there is, corresponding to the phenomenon of conical refraction just considered, a reciprocal law which refers to the normal direction and which is founded on the circumstance that certain singular rays have not two but an infinite number of different wave-normals corresponding to them. We have then only to endeavour to produce a ray whose direction coincides with a secondary optic axis. For this purpose we may very well use the same arragonite plate as in the experiment described in the preceding section, but in the following way.

We place a point-source of light O at the surface of the crystal, say by covering the surface by an opaque screen having a fine hole O and then illuminating the screen from all directions. Rays from O will then enter the crystal in all directions. If we now also cover the opposite surface of the plate likewise by an opaque screen having a hole at P, then of all the entrant rays only a single one, namely the one OP, can emerge from the crystal. As we know, there are in general two normals, that is, two wave-planes corresponding to this ray, which make certain angles with them and pass out into the air according to a definite law of refraction which we shall learn more about in the next chapter. Thus at P we obtain two rays of light emerging in different directions.

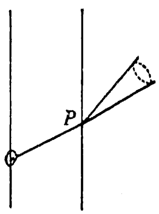

FIG. 20.

But in the special case where the direction OP coincides with a secondary optic axis of the crystal, there are an infinite number of differently directed normals and wave-planes belonging to the ray OP in the crystal, and since each of them is refracted in a particular way on emerging from the crystal, the light from P propagates itself conically in an infinite number of different directions into the air (Fig. 20), a phenomenon which is known as "external conical refraction" because here the resolution of the light occurs on leaving the crystal.

To find the appropriate direction P we must, of course, arrange to be able to move one of the screens, and must move it until the straight line OP connecting the two holes coincides with a secondary optical axis.

Of course in the case of conical refraction, both internal and external, each of the rays of the cylinder or cone that emerges into the air has a perfectly definite plane of vibration and polarization, which may be obtained from the well-known laws of propagation and refraction.

CHAPTER IV

OBLIQUE INCIDENCE

§ 68. The laws governing the passage of a plane wave of light from an isotropic body into a crystal, or the converse, may be derived by the same method as was used in the first part of this volume: we first form the expressions for plane waves in the interior of the two bodies in contact and then set up the boundary conditions for the surface of separation which express the continuity of the tangential components of the electric and magnetic intensity of field. We get from this, if the incident wave is arbitrarily given, definite expressions for the refracted and the reflected wave.

Since, in order to be able to retain the relationships that have hitherto been used, we have again taken the principal axes of the crystal as co-ordinate axes, we cannot here, as we could in the case of isotropic bodies, take the normal of the incident wave as the x-axis; rather, we shall call the direction cosines of this normal α_0, β_0, γ_0 and the velocity of propagation in the first body, which is isotropic, q_0. Then, as in (8), an incident wave of definite polarization will be characterized by its wave-function (magnetic intensity of field):

$$f\left(t - \frac{\alpha_0 x + \beta_0 y + \gamma_0 z}{q_0}\right) \quad . \quad . \quad . \quad (297)$$

where f represents an arbitrarily given function of a single argument. Let the normal of the plane of separation, the incident normal, have the direction cosines u, v, w. We wish to find the wave-functions of the refracted wave:

$$f_1\left(t - \frac{\alpha x + \beta y + \gamma z}{q}\right) \quad . \quad . \quad . \quad (298)$$

and the reflected wave:

$$f'\left(t - \frac{\alpha'_0 x + \beta'_0 y + \gamma'_0 z}{q_0}\right) \quad . \quad . \quad . \quad (299)$$

As in the case of isotropic bodies so here the problem falls into two parts. Firstly we have to determine the directions of the refracted and the reflected wave; secondly we have to determine the wave-functions f_1 and f'. Both these objects are accomplished by setting up the boundary conditions which express the equality of the electric and magnetic intensity of field at both sides of the plane of separation:

$$ux + vy + wz = 0 \quad . \quad . \quad . \quad . \quad (300)$$

As in (15) and (16), this again gives us two equations which are linear and homogeneous in the quantities f, f_1, f', from which we find that f_1 and f' are proportional to f.

§ 69. Let us first consider the directions of the refracted and the reflected wave. Since the proportionality of f_1 and f' with f is to hold for all times t and for all points of the boundary surface (300), therefore, just as in § 7, the arguments of these functions must be equal to each other; that is, at the boundary surface we must have:

$$\frac{\alpha_0 x + \beta_0 y + \gamma_0 z}{q_0} = \frac{\alpha x + \beta y + \gamma z}{q} = \frac{\alpha'_0 x + \beta'_0 y + \gamma'_0 z}{q_0} \quad (301)$$

The *first* equation leads to:

$$\left(\frac{\alpha}{q} - \frac{\alpha_0}{q_0}\right)x + \left(\frac{\beta}{q} - \frac{\beta_0}{q_0}\right)y + \left(\frac{\gamma}{q} - \frac{\gamma_0}{q_0}\right)z = 0 \quad . \quad (302)$$

If this were to hold for any arbitrary values of x, y, z the bracketed quantities would all have to vanish. But x, y, z are connected together by the condition (300). So we may also proceed by expressing z in terms of x and y, substituting in (302) and then setting the coefficients of x and y both equal to zero. It is more expedient to use Lagrange's method of undetermined multipliers as in I, § 97. This consists in multiplying the expression (300)

by a certain constant λ, adding the result to (302), and then setting the coefficients of x y, z individually equal to zero. This gives us the three equations:

$$\left.\begin{aligned} \frac{\alpha}{q} - \frac{\alpha_0}{q_0} + \lambda u &= 0 \\ \frac{\beta}{q} - \frac{\beta_0}{q_0} + \lambda v &= 0 \\ \frac{\gamma}{q} - \frac{\gamma_0}{q_0} + \lambda w &= 0 \end{aligned}\right\} \quad . \quad . \quad . \quad (303)$$

These equations allow of a simple geometrical interpretation. In the first place the three directions represented by the direction cosines all lie in a plane, that is, the normal of the refracted wave lies in the plane of incidence formed by the normal of the incident wave and the normal to the surface; and secondly, if we multiply the individual equations with the corresponding direction cosines of the line of intersection of the incident plane and the boundary plane and then add, we get:

$$\frac{\sin\theta}{q} - \frac{\sin\theta_0}{q_0} = 0 \quad . \quad . \quad . \quad . \quad (304)$$

where θ and θ_0 denote the angles between the two wave-normals and the normal to the plane. Hence Snell's law of refraction also holds for refraction at a crystal surface so far as it refers to the normal of the refracted wave. But here it has not the same significance as in the case of isotropic bodies, because the velocity of propagation q is not known and itself depends on the desired angle of refraction θ. Hence in order to calculate the direction of the refracted wave we must use the relationship between q and α, β, γ expressed by (264). This is done most simply and intelligibly by introducing the abbreviations:

$$\left.\begin{aligned} \frac{\alpha}{q} = x, \ \frac{\beta}{q} = y, \ \frac{\gamma}{q} = z \\ \frac{1}{q^2} = x^2 + y^2 + z^2 = r^2 \end{aligned}\right\} \quad . \quad . \quad . \quad (305)$$

where, of course, x, y, z now have a different meaning from before. For then it follows from (303) that:

$$\left(x-\frac{\alpha_0}{q_0}\right):\left(y-\frac{\beta_0}{q_0}\right):\left(z-\frac{\gamma_0}{q_0}\right)=u:v:w \quad . \quad . \quad (306)$$

and from (264):

$$\frac{x^2}{a^2r^2-1}+\frac{y^2}{b^2r^2-1}+\frac{z^2}{c^2r^2-1}=0 \quad . \quad . \quad (307)$$

The first equation states that the point P represented by the co-ordinates x, y, z lies on the straight line which is parallel to the normal u, v, w to the plane of the crystal and which passes through the point A whose co-ordinates are $\frac{\alpha_0}{q_0}, \frac{\beta_0}{q_0}, \frac{\gamma_0}{q_0}$. Here OA and OP (Fig. 21) denote the directions of the normals of the incident and the refracted wave and the reciprocal values of the corresponding velocities of propagation q_0 and q. The second equation states that the point P lies on the surface (307). Comparison with (284) shows that this surface may be regarded as the optical wave-surface of a hypothetical crystal whose principal velocities of propagation are $\frac{1}{a}, \frac{1}{b}, \frac{1}{c}$. This surface is also called the "index surface" of the real crystal whose principal velocities of propagation are a, b, c, because the principal indices of refraction n_1, n_2, n_3 are, by (255), inversely proportional to the quantities a, b, c.

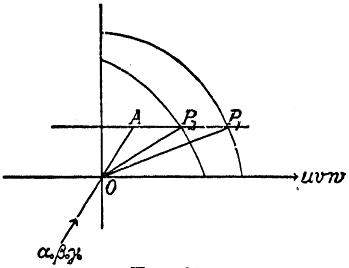

FIG. 21.

The index surface bears the same relationship to the wave-surface as the ellipsoid whose semi-axes are a, b, c (§ 60) bears to the Cauchy ellipsoid of polarization whose semi-axes are $\frac{1}{a}, \frac{1}{b}, \frac{1}{c}$ (§ 58). Hence all the consequences

that follow from the general law of reciprocity also apply to the properties of the index surface. For example, the primary optic axes of the index surface are at the same time the secondary optic axes of the wave-surface and vice versa.

§ 70. The preceding theorems give a simple geometrical method for finding the direction and the velocity of propagation of the refracted wave. Through the central point O of the index surface of the crystal draw a plane parallel to the plane of incidence and take it as the plane of the diagram (Fig. 21). This plane contains the normal $\alpha_0\beta_0\gamma_0$ of the incident wave and the normal of incidence uvw. The intersection with the two shells of the index plane is denoted in the figure by two curves, one of which encloses the other. Now draw the distance $OA = \frac{1}{q_0}$ in the direction $\alpha_0\beta_0\gamma_0$ and through A draw the parallel to the normal of incidence uvw. Every point of intersection P_1 or P_2 with the index surface then gives, when joined to O (that is, the line OP_1, or OP_2), the direction $\alpha\beta\gamma$ and the reciprocal velocity of propagation $\frac{1}{q}$ of the normal of the refracted wave. So in general when the incident wave enters the crystal it resolves into two different refracted waves. Corresponding to each of these two wave-normals there is (1) a definite ray whose direction is represented by the normal to the index plane at the point P, and so in general leaves the plane of incidence and (2) a definite plane of vibration which passes through the normal and the ray, so that the direction of the electric and magnetic intensity of field and of the induction is determined.

If q_0 is not too great compared with the principal velocities of propagation a, b, c it can happen that the point A lies outside the inner or even the outer shell of the index surface. In this case the point P_2 or both points P_1 and P_2 possibly become imaginary, and the conditions for total reflection obtain.

The method of construction just described is graphically verified if we apply it to the refraction of light by an isotropic body, which may always be regarded as a special case of a crystal. In the case of an isotropic body the two curves of intersection of the index-surface coincide with the circle of radius $\frac{1}{q}$, and from the relationships between the angles θ_0 and θ, which the directions OA and OP form with the normal of incidence uvw, and the lengths $\frac{1}{q_0}$ and $\frac{1}{q}$ of these distances we obtain Snell's law of refraction $\sin\theta_0 : \sin\theta = q_0 : q$.

§ 70*a*. For the *reflected* wave the second equation (301) gives:

$$(\alpha'_0 - \alpha_0)x + (\beta'_0 - \beta_0)y + (\gamma'_0 - \gamma_0)z = 0$$

and it follows from this, by reasoning fully analogous to that given above, that the normal of the reflected wave lies in the plane of incidence, and that its angle θ'_0 with the incident normal is determined by:

$$\sin\theta'_0 = \sin\theta_0.$$

Hence:

$$\theta'_0 = \pi - \theta_0.$$

That is, the angle of reflection is equal to the angle of incidence, just as in the case of isotropic bodies.

§ 71. In the same way as we have here treated the refraction and reflection of a wave that emerges from an isotropic body and falls on a crystal we may also treat the converse case where a wave which advances within a crystal impinges on the surface of an isotropic body. We then obtain by the same method the corresponding laws of refraction and reflection, which distinguish themselves in a characteristic way from those obtained earlier. In general, for example, there are two different waves which are reflected back into the crystal from the surface, whereas only a single wave which can be determined directly from Snell's law escapes into the isotropic medium.

§ 72. We shall select for treatment only a few of the many applications which can be made of the laws which we have derived for the reflection and refraction of crystals. To simplify matters as much as possible we shall restrict ourselves to *uniaxial positive* crystals, so that by § 59 :

$$a = b > c \quad . \quad . \quad . \quad . \quad . \quad (308)$$

For this the equation (264) becomes :

$$\{(\alpha^2 + \beta^2)(c^2 - q^2) + \gamma^2(a^2 - q^2)\} \cdot (a^2 - q^2) = 0$$

whose roots are :

$$\left.\begin{aligned} q_1 &= a \\ q_2{}^2 &= (\alpha^2 + \beta^2)c^2 + \gamma^2 a^2 \end{aligned}\right\} \quad . \quad . \quad . \quad (309)$$

The wave whose constant velocity of propagation is $a = q_1$ is called the "ordinary" wave and that with the variable velocity of propagation q_2 is called the "extraordinary" wave.

If we call the angle which the wave-normal makes with the optic axis, the z-axis, θ, that is, $\gamma = \cos\theta$, and if, for simplicity, we omit the suffix 2 in q_2, then we get for the velocity of propagation of the extraordinary wave, from (309) :

$$q^2 = a^2\cos^2\theta + c^2\sin^2\theta \quad . \quad . \quad . \quad (310)$$

The equation (285) of the optical wave-surface becomes, for $a = b$:

$$(a^2x^2 + a^2y^2 + c^2z^2 - a^2c^2) \cdot (x^2 + y^2 + z^2 - a^2) = 0 \quad (311)$$

So the surface resolves into the elongated ellipsoid of rotation whose semi-axes are c, c, a :

$$\frac{x^2}{c^2} + \frac{y^2}{c^2} + \frac{z^2}{a^2} = 1 \quad . \quad . \quad . \quad . \quad (312)$$

and into the sphere of radius a, where :

$$x^2 + y^2 + z^2 = a^2$$

which touches the ellipsoid at the ends of its axis.

The ellipsoid forms the inner shell, the sphere the

outer shell. The primary and the secondary optic axes coincide with the z-axis.

In Fig. 22 the two wave-planes are drawn which correspond to a definite wave-normal OQ. They touch the wave-surface at the point P of the internal and Q of the external shell. The corresponding rays are OP and OQ. The vibration-plane of the extraordinary ray OP coincides with the plane of the figure, namely the principal section of the crystal that contains the optic axis; the vibration-plane of the ordinary ray OQ is perpendicular to the latter. Accordingly the electric induction at P also lies in the plane of the diagram, whereas that at Q is at right angles to it (cf. Figs. 14 and 15).

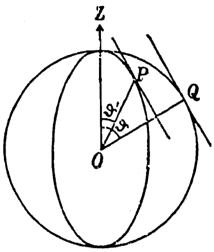

FIG. 22.

The angle which the ordinary ray OQ makes with the z-axis is the same as that between the wave-normals, namely θ; that which the extraordinary ray OP makes with the z-axis is θ'. The value of θ' can easily be obtained by reflecting that in Fig. 22 the angle θ denotes the direction of the normal to the ellipse at the extremity of the diameter OP drawn in the direction θ'. Thus:

$$\sin\theta : \cos\theta = \frac{\sin\theta'}{c^2} : \frac{\cos\theta'}{a^2} \quad . \quad . \quad . \quad (313)$$

From this we obtain:

$$\cos\theta' = \frac{a^2\cos\theta}{\sqrt{a^4\cos^2\theta + c^4\sin^2\theta}} \quad . \quad . \quad (314)$$

For the angle δ between the ray and the normal of the extraordinary wave we get:

$$\cos\delta = \cos(\theta - \theta') = \frac{a^2\cos^2\theta + c^2\sin^2\theta}{\sqrt{a^4\cos^2\theta + c^4\sin^2\theta}} \quad . \quad (315)$$

and for the velocity of propagation measured in the direction of the ray:

$$q' = \frac{q}{\cos\delta} = \sqrt{\frac{a^4\cos^2\theta + c^4\sin^2\theta}{a^2\cos^2\theta + c^2\sin^2\theta}} \quad . \quad . \quad (316)$$

In calculating these quantities we may of course use the relationships which arise from the general law of reciprocity (§ 60). For example, from equation (310) we may immediately derive the following relationship :

$$\frac{1}{q'^2} = \frac{\cos^2\theta'}{a^2} + \frac{\sin^2\theta'}{c^2} \quad . \quad . \quad . \quad . \quad (317)$$

which may easily be verified subsequently.

§ 73. We shall now consider the passage of a plane monochromatic wave through a plane-parallel plate of a uniaxial positive crystal which is cut perpendicularly to the optic axis. Let the angle of incidence be θ_0, and let the plane of incidence, which at the same time represents a principal section of the crystal, be taken as the plane of the diagram in Fig. 23, in which the z-axis represents the optic axis and the incident ray is imagined as passing from below on the left through the plate up towards the right.

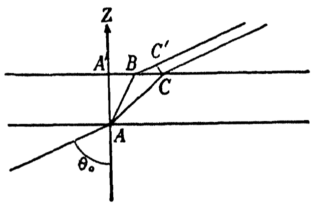

FIG. 23.

Since we shall assume the wave, like the plate, to be unlimited laterally we need not take the ray-direction into special consideration and may restrict ourselves to considering the wave-normals.

A ray of natural light which falls from below on the point A of the plate there resolves according to Snell's law into an ordinary wave :

$$\frac{\sin\theta_0}{\sin\theta_1} = \frac{q_0}{q_1} = \frac{q_0}{a} \quad . \quad . \quad . \quad . \quad . \quad (318)$$

and an extraordinary wave :

$$\frac{\sin\theta_0}{\sin\theta_2} = \frac{q_0}{q_2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (319)$$

where q_2 and θ_2 are related to each other as in (310). Each of these two waves traverses the crystal and in passing out through the opposite face, which contains the point A, each yields a ray which is directed upwards

towards the right which is parallel to the ray which extends from below at A, since the second refraction occurs in exactly the opposite direction sense to the first. Hence the two waves subsequently reunite again to form a single wave which is parallel to the incident wave. But since the path measured along the wave-normal is different for the two waves the phase-difference between them when they leave the plate is different from that on entering. We proceed to calculate this phase-displacement Δ which arises from transmission through the plate.

The normal of the ordinary wave in the crystal is represented in Fig. 23 by the distance AC, and the angle $CAA' = \theta_1$. At the point C the second refraction occurs which restores the original direction which made an angle θ_0 with the z-axis.

The normal of the extraordinary wave in the crystal is represented by the distance AB, the angle BAA' being equal to θ_2. At the point B the ray again escapes into the air in the direction θ_0. The wave-plane of the transmitted light is then CC' which is perpendicular to BC', and the wave-normal is BC'. The desired phase-displacement Δ is obtained if we fix our attention on two definite wave-planes, the one before its entrance into the plate, say that which passes through A, the other after its emergence from the plate, say that through C, and compare the changes which have occurred in the wave-function $\omega\left(t - \frac{n}{q}\right)$ in the case of each of the two waves in the intervening space between the two wave-planes. In this intervening space the normal of the ordinary wave is the straight line AC, that of the extraordinary wave is the refracted straight line ABC', of which the part AB is traversed with the velocity q_2, and the part BC' with the velocity q_0, no sudden change of phase occurring in the process of refraction. From this we get as the required phase-displacement:

$$\Delta = \omega\left(\frac{AB}{q_2} + \frac{BC'}{q_0} - \frac{AC}{a}\right) \quad . \quad . \quad . \quad (320)$$

The distances AB, BC', AC as well as the velocity of propagation q_2 are obtained from the angle of incidence θ_0. For simplicity we assume θ_0 to be small and restrict our attention to terms involving θ_0^2. It then follows from (318) that:

$$\theta_1 = \frac{a}{q_0} \cdot \theta_0, \quad \cos \theta_1 = 1 - \frac{a^2\theta_0^2}{2q_0^2}$$

and from (319) that:

$$\theta_2 = \frac{q_2}{q_0} \cdot \theta_0$$

and by combining these results with (310) we get:

$$q_2^2 = a^2 - (a^2 - c^2)\theta_2^2$$

$$q_2 = a\left(1 - \frac{a^2 - c^2}{2q_0^2}\theta_0^2\right)$$

$$\theta_2 = \frac{a}{q_0}\theta_0 = \theta_1.$$

Further, if D is the thickness of the plate:

$$AB = \frac{D}{\cos \theta_2} = D\left(1 + \frac{a^2\theta_0^2}{2q_0^2}\right)$$

$$AC = \frac{D}{\cos \theta_1} = AB$$

whereas the distance:

$$BC' = BC \cdot \theta_0$$

is of a smaller order than $D \cdot \theta_0^2$ and may therefore be neglected.

The substitution of all these values in (320) leads to:

$$\Delta = \omega \cdot AB \cdot \left(\frac{1}{q_2} - \frac{1}{a}\right) = \frac{\omega D}{2a} \cdot \frac{a^2 - c^2}{q_0^2} \cdot \theta_0^2$$

or, if we introduce the principal coefficients of refraction n_1 and n_2 and the wave-length λ_0 in air we get:

$$\Delta = \frac{\pi D n_1}{\lambda_0}\left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right) \cdot \theta_0^2 \quad . \quad . \quad . \quad (321)$$

The extraordinary wave has its vibration plane and hence also its electric induction in the principal section, that is, in the plane of incidence, the ordinary wave being perpendicular to it. Hence the extraordinary wave is given by that component of the incident ray which vibrates in the plane of incidence and the ordinary wave is given by the component which is perpendicular to it.

If the incident light is linearly polarized, say by a Nicol prism used as a polarizer, whose plane of vibration makes the angle ϕ with the plane of incidence, the phase-difference between the two waves on entering the plate is zero and their amplitudes are in the ratio $\cos\phi : \sin\phi$, if we discard losses due to reflection. On leaving the plate the phase-difference is, as we calculated, Δ, so that the emergent light is in general elliptically polarized. If we subsequently allow the light to pass through a second Nicol prism, acting as an analyser, whose plane of vibration makes the angle ψ with the plane of incidence, we obtain just as in (295) the intensity of the transmitted light referred to the light delivered by the polarizer as unity:

$$J = \cos^2(\phi - \psi) - \sin 2\phi \,.\, \sin 2\psi \,.\, \sin^2\frac{\Delta}{2} \quad . \qquad (322)$$

where Δ is now given by (321).

If the two Nicols are crossed, then we have as in (296):

$$J = \sin^2 2\phi \,.\, \sin^2\frac{\Delta}{2} \quad . \quad . \quad . \quad . \qquad (323)$$

If, finally, the light passes through a focusing lens which is parallel to the crystal plate the light rays meet at a single point of the focal plane of the lens, whose position is denoted by the angle θ_0, which is proportional to the distance of the point from the axial point, that is, the focus of the rays that pass perpendicularly through the plate without being refracted.

From this we can derive the phenomena that occur when a whole beam of nearly normally incident rays pass through a plane-parallel plate of a uniaxial crystal cut

perpendicularly to the optic axis and situated between two crossed Nicols. Here we have an infinite number of planes of incidence adjacent to one another, each corresponding to a different value of ϕ. Hence the intensity of light J, besides depending on θ_0, also depends on the azimuthal angle ϕ of the plane of incidence, and we obtain in the focal plane of the lens a black cross $\left(\phi = 0, \phi = \frac{\pi}{2}\right)$ with concentric black rings superposed on it, whose radii are given by those values of Δ which are integral multiples of 2π. For other positions of the Nicol prism we obtain the value of J directly from (322).

§ 74. So far we have dealt only with the direction and the velocity of propagation of the waves refracted and reflected by a crystal.

If we now revert to the remarks made in §68 and inquire into the values of the wave-functions we must take into account, just as in the case of isotropic bodies, the special form of the boundary conditions. But the problem will be considerably more complicated than for isotropic bodies because here we are no longer able to treat the refraction and the reflection of the component vibrating in the plane of incidence completely separately from the component that vibrates perpendicularly to the plane of incidence. In general, rather, each of these two components will undergo double refraction when it encounters the crystal.

Nevertheless a moment's consideration will show that there is a way of separating the problem into two independent parts and so to simplify the calculation considerably. Only, in doing so, we must take care not to start with a definite component of the wave incident from the isotropic body, but must fix our attention primarily on one of the two waves advancing in the crystal. For if, with the help of the theorems of the last section, the directions of propagation and the polarization, as well as the velocities of propagation of the two refracted waves, are completely determined we may propose the

question: of what nature must the incident wave be in order that only one of the two refracted waves is formed?

So we now consider as known the direction of the normal n, of the electric and magnetic intensity of field and of the induction, as well as of the velocity of propagation q of one of the two refracted waves, and we determine the corresponding incident and reflected wave. This is accomplished by setting up the boundary conditions, which express the continuity of the tangential components of the field-strengths.

If we denote the magnetic intensity of field of the *refracted* wave by:

$$H = f_1\left(t - \frac{n}{q}\right) \quad . \quad . \quad . \quad . \quad (324)$$

then the electric intensity of field of the wave is, by (257):

$$E = \frac{q}{c_0 \cos \delta} \cdot f_1\left(t - \frac{n}{q}\right) \quad . \quad . \quad . \quad (325)$$

The direction of the electric intensity of field, which we also take as known, may be conveniently expressed by means of its direction cosines, ξ η, ζ, referred to the three following mutually-perpendicular directions: the normal of incidence (ξ), the line of intersection of the incident plane with the boundary face (η) and the normal of the plane of incidence (ζ) (cf. Fig. 1).

Then, of the two components of the electric intensity of field which are tangential to the boundary surface, that in the plane of incidence is

$$\frac{q}{c_0 \cos \delta} \cdot f_1 \cdot \eta \quad . \quad . \quad . \quad . \quad . \quad . \quad (326)$$

and that perpendicular to the plane of incidence is:

$$\frac{q}{c_0 \cos \delta} \cdot f_1 \cdot \zeta \quad . \quad . \quad . \quad . \quad . \quad . \quad (327)$$

The direction of the magnetic intensity of field is perpendicular firstly to the electric intensity of field which

has the direction cosines ξ, η, ζ, and secondly to the wave-normal which lies in the plane of incidence which has the direction cosines $\cos\theta$, $\sin\theta, 0$. Hence the direction cosines of the magnetic intensity are:

$$\frac{\zeta\sin\theta}{\cos\delta},\quad \frac{-\zeta\cos\theta}{\cos\delta},\quad \frac{\eta\cos\theta-\xi\sin\theta}{\cos\delta}. \quad . \qquad (328)$$

The correct signs are obtained by considering a special case, such as that, for example, for which $\delta=0$, $\theta=0$, $\eta=1$. Hence, by (324) and (328), of the two components of the magnetic field-strength which are tangential to the boundary surface that which lies in the plane of incidence is:

$$-f_1\cdot\frac{\zeta\cos\theta}{\cos\delta}. \quad . \quad . \quad . \quad . \quad . \qquad (329)$$

and that perpendicular to the plane of incidence is:

$$f_1\frac{\eta\cos\theta-\xi\sin\theta}{\cos\delta} \quad . \quad . \qquad (330)$$

We next form the corresponding quantities for the *incident* wave, which we, of course, also assume to be linearly polarized. If we denote its magnetic field-strength by f, then by (8) and (7) its electric field-strength is $f\cdot\frac{q_0}{c_0}$. If the vibration-plane makes the angle ϕ with the plane of incidence, then of the two components of the electric field-strength which are tangential to the boundary surface that in the incident plane is:

$$f\cdot\frac{q_0}{c_0}\cdot\cos\phi\cdot\cos\theta_0 \quad . \quad . \quad . \quad . \quad . \qquad (331)$$

and that perpendicular to the incident plane is:

$$f\frac{q_0}{c_0}\sin\phi \quad . \quad . \quad . \quad . \quad . \qquad (332)$$

while the corresponding quantities for the magnetic field-strength are:

$$-f\sin\phi\cos\theta_0 \quad . \quad . \quad . \quad . \quad . \qquad (333)$$

and :

$$f \,.\, \cos \phi \quad . \quad . \quad . \quad . \quad . \quad . \qquad (334)$$

Finally, for the *reflected* wave we obtain fully analogous expressions, which we distinguish, as before, by accenting them, noting that $q'_0 = q_0$ and $\theta'_0 = \pi - \theta_0$. For the electric intensity of field we obtain in this way the components :

$$-f' \cdot \frac{q_0}{c_0} \cos \phi' \cos \theta_0 \quad . \quad . \quad . \quad . \quad . \qquad (335)$$

and :

$$f' \frac{q_0}{c_0} \sin \phi' \quad . \quad . \quad . \quad . \quad . \quad . \qquad (336)$$

For the magnetic intensity of field we get the components :

$$f' \sin \phi' \cos \theta_0 \quad . \quad . \quad . \quad . \quad . \quad . \qquad (337)$$

and :

$$f' \cos \phi' \quad . \quad . \quad . \quad . \quad . \quad . \qquad (338)$$

At the boundary surface the tangential components of all the field-strengths are continuous, that is, the sum of each pair of field-strengths, of the same kind, of the incident and reflected wave is equal to the corresponding field-strength of the refracted wave. This gives us four boundary conditions; namely, firstly, for the electric field-strength in the incident plane, by (331), (335), (326):

$$fq_0 \cos \phi \cos \theta_0 - f'q_0 \cos \phi' \cos \theta_0 = \frac{q}{\cos \delta} f_1 \eta \qquad (339)$$

secondly, for the electric field-strength perpendicular to the plane of incidence, by (332), (336), (327) :

$$fq_0 \sin \phi + f'q_0 \sin \phi' = \frac{q}{\cos \delta} f_1 \zeta \quad . \quad . \qquad (340)$$

thirdly, for the magnetic intensity of field in the plane of incidence, by (333), 337), (329) :

$$-f \sin \phi \cos \theta_0 + f' \sin \phi' \cos \theta_0 = -f_1 \cdot \frac{\zeta \cos \theta}{\cos \delta} \quad . \qquad (341)$$

and fourthly, for the magnetic intensity of field perpendicular to the plane of incidence, by (324), (338), (330):

$$f \cos \phi + f' \cos \phi' = f_1 \frac{\eta \cos \theta - \xi \sin \theta}{\cos \delta} \quad . \quad (342)$$

If, as in § 7, we set:

$$f' = \mu f \text{ and } f_1 = \mu_1 f \quad . \quad . \quad . \quad (343)$$

the four equations may be satisfied by definite values of ϕ, ϕ', μ and μ_1; that is, there is in fact an incident wave of definite form and polarization which, besides producing a reflected wave of a definite but different form and different polarization, only calls up in the crystal the refracted wave here introduced. In view of (304) the values in question come out as:

$$\tan \phi = \frac{\zeta \sin (\theta_0 + \theta)}{\eta(\sin \theta_0 \cos \theta_0 \cos \theta + \sin \theta) - \xi \sin \theta_0 \cos \theta_0 \sin \theta} \quad (344)$$

$$\tan \phi' = \frac{\zeta \sin (\theta_0 - \theta)}{\xi \sin \theta_0 \cos \theta_0 \sin \theta - \eta(\sin \theta_0 \cos \theta_0 \cos \theta - \sin \theta)} \quad (345)$$

$$\mu = - \frac{\sin \phi}{\sin \phi'} \cdot \frac{\sin (\theta_0 - \theta)}{\sin (\theta_0 + \theta)} \quad . \quad . \quad . \quad (346)$$

$$\mu_1 = \frac{\sin \phi \cdot \cos \delta \cdot \sin 2\theta_0}{\zeta \cdot \sin (\theta_0 + \theta)} \quad . \quad . \quad . \quad (347)$$

For the special case where the electrical intensity of field of the refracted wave is perpendicular to the wave-normal and lies in the plane of incidence ($\xi = - \sin \theta$, $\eta = \cos \theta$, $\zeta = 0$, $\delta = 0$) the values of μ and μ_1 assume the similarly named values (23) and (24) (angle of incidence θ_0, angle of refraction θ), whereas the angles ϕ and ϕ' both vanish.

§ 75. If we next enquire how the incident wave must be constituted, for the same angle of incidence θ_0, in order that only the second of the two refracted waves may come about, we may of course find the answer in the same way by starting out from the wave-function of the second refracted wave: $g_1\left(t - \frac{n_2}{q_2}\right)$. The correspond-

ing calculations then again lead to a definite incident wave g which vibrates in the azimuth ϕ_2 and a definite reflected wave g' which vibrates in the azimuth ϕ'_2. If, as in § 7, we then set

$$g' = \sigma g \quad \text{and} \quad g_1 = \sigma_1 g$$

the last four equations of the preceding section give the corresponding values of ϕ_2, ϕ'_2, σ and σ_1, if we substitute for θ, δ, ξ, η, ζ the quantities calculated for the second refracted wave.

For the special case where the electric intensity of field of the refracted wave is perpendicular both to the wave-normal and to the plane of incidence ($\xi = 0$, $\eta = 0$, $\zeta = 1$, $\delta = 0$) the values of σ and σ_1 become equal to the similarly named values (23) and (24), while the angles ϕ_2 and ϕ'_2 both become equal to $\frac{\pi}{2}$.

From this we may now obtain the solution of the general problem—namely that of finding, for an arbitrarily given incident wave, the form of the two refracted waves as well as the form and the polarization of the one reflected wave. We resolve the incident wave into the two wave-components f and g, each of which produces independently of the other a refracted and a reflected wave according to a definite law. Only we may not here, as was possible in the case of isotropic bodies, resolve the vibrations in the azimuths 0 and $\frac{\pi}{2}$—that is, in and perpendicular to the plane of incidence, but rather in the azimuths ϕ_1 and ϕ_2, as given by (344), if we substitute for θ, ξ, η, ζ the values calculated for the two refracted waves from the angle of incidence θ_0 according to § 69. Then $\mu_1 . f$ and $\sigma_1 . g$ are the wave-functions of the two refracted waves, whereas $\mu . f$ and $\sigma . g$, which have the same normal, combine to form the reflected wave according to their azimuths ϕ'_1 and ϕ'_2. If the incident light is linearly polarized, then so is the reflected light; if it is natural light, then the reflected light is partially polarized.

It is obvious that the results thus obtained now allow us to treat all the special applications previously made for isotropic bodies also for crystals. For example, we may calculate the coefficients of reflection and transmission (ρ and $1-\rho$), normal incidence ($\theta_0 = 0$), grazing incidence $\left(\theta_0 = \frac{\pi}{2}\right)$, angle of polarization ($\rho = 0$), total reflection (θ imaginary). It would lead us too far to discuss all these questions in detail here.

PART THREE

DISPERSION OF ISOTROPIC BODIES

CHAPTER I

FUNDAMENTAL EQUATIONS

§ 76. REVERTING from now on to the consideration of isotropic bodies we shall next occupy ourselves with a phenomenon which intrudes itself to a greater or a lesser degree and which we have been compelled to ignore up to the present because it has no place at all in the original theory of Maxwell—namely *dispersion*. We encountered this peculiar circumstance quite early, in § 9, and we have already envisaged the only escape from the difficulty in which it involves the theory: this consists in letting fall the assumption that space is filled absolutely continuously by matter and introducing instead an atomistic point of view.

In taking this step we essentially discard the point of view to which we have hitherto clung without exception in all our discussions and enter a field in which, to be able to make progress at all, we are from the very outset at the mercy of more or less arbitrary hypotheses. This difficulty is inherent in the nature of the problem and may be mitigated to a certain extent by assuming the atoms or, expressed more generally, the smallest elementary bricks (*Bausteine*) of matter to be extremely tiny and extremely numerous. While this assumption provides us with the possibility, on the one hand, of attributing to the material body which is composed of atoms new kinds of properties which have not been expressible in the theory hitherto developed, the body yet behaves practically, provided that the phenomena in question do not occur in dimensions of too small an order of magnitude, as if it were perfectly homogeneous. A good

example of this is given by the propagation of sound in a body constituted of atoms, for which, provided the wave-length is not too small, exactly the same laws hold as in an absolutely homogeneous body of corresponding density and elasticity. Regarded from this point of view the whole theory hitherto discussed appears as a limiting theory which is valid for comparatively slow and " coarse-grained " events, and our present task is to extend the range of validity of the theory a step further towards the side of more rapid and " finer-grained " events.

§ 77. To preserve connection with what has gone before we retain, as in crystal optics, the fundamental equations (1), (2) and (3) and only introduce a more general relationship in place of equation (4), which expresses the proportionality between the electric induction $\boldsymbol{D}$ and the electric intensity of field $\boldsymbol{E}$. Such a relationship is suggested to a certain extent by a graphical physical interpretation which can be given, by III, § 26, to the difference between induction and field strength. For if we set, by III (141):

$$\frac{\boldsymbol{D} - \boldsymbol{E}}{4\pi} = \boldsymbol{M} \quad . \quad . \quad . \quad . \quad . \qquad (348)$$

$\boldsymbol{M}$ may be regarded as the electric moment of the unit of volume, that is, as the sum of the moments of all the infinitely small electric dipoles which are to be supposed contained in the volume 1 (taken sufficiently small).

The advance which we now intend to make beyond this view consists in ascribing individually to each of these dipoles a real physical existence, that is, we no longer assume an infinite number of infinitely small dipoles in immediate contact with one another, to enquire into the individual size of which has no meaning, but rather we assume that their number, size and distances from one another are real and finite. The most essential feature of this hypothesis is that now there is only a single medium which is continuous in the true sense, namely, the vacuum in which all electromagnetic field

effects occur, whereas the material bodies share only in a secondary way in these processes, in that their electrically charged components, the ions and electrons, fly about individually in the vacuum. This endows the critical velocity c, the velocity of propagation of light in a vacuum, with a much deeper meaning than it had in the original theory, where it was only one velocity of propagation among many.

§ 78. If we now enquire what general relationship may take the place of equation (4) if we adopt the new view suggested above, we must first reflect that by equation (4) the electric moment of all the dipoles contained in volume 1 is proportional to the simultaneous field-strength. As we now wish to ascribe a real existence to the dipoles, we shall also have to give them a certain independence; that is, we shall not demand that their electrical moment owing to some intimate coupling shall instantaneously participate in all the fluctuations of the electric field that acts on them, but rather we shall generalize the law for the relationship between the electrical moment and the exciting electrical intensity of field by distinguishing between the fluctuations of the electrical moment of the dipoles and the fluctuations of the exciting electrical intensity of field. The former will, indeed, be determined by the latter, not by means of a simple relationship of proportionality, but by means of a special equation of vibration in which also the individual properties of the dipoles, in particular their inertia, play a part. There are various possibilities for choosing the form of the equation of vibration. For example, we may assume the dipoles to be rigid and capable of rotation and having constant electrical moments. Their moments of inertia then enter into the equation of vibration and the electrically neutral state is characterized by the completely chaotic arrangement of the dipoles.

The phenomena of optical dispersion here under discussion can be best accounted for, at least to a first order of approximation, if we represent the vibration of

a dipole by means of the equation for the forced oscillations of a point-mass in carrying an invariable electric charge e and oscillating about a definite position of equilibrium which has an equal and opposite charge. In vectorial form this equation runs (II, 203):

$$m\ddot{\boldsymbol{r}} + m\omega_0^2\boldsymbol{r} = e\,.\,\boldsymbol{E}' \quad . \quad . \quad . \quad . \qquad (349)$$

Here ω_0 denotes the proper frequency of the oscillating dipole, $\boldsymbol{r}$ its distance from the position of equilibrium, and hence, by III, § 26, $e\,.\,\boldsymbol{r}$ its electric moment, and, if there are N such dipoles in unit volume:

$$N\,.\,e\,.\,\boldsymbol{r} = \boldsymbol{M} \quad . \quad . \quad . \quad . \quad . \qquad (350)$$

is the electric moment per unit of volume.

The exciting field-strength $\boldsymbol{E}'$ must be carefully distinguished from the total field-strength $\boldsymbol{E}$. For the former arises only from the electric charges which are situated outside the dipole in question, since a dipole does not excite itself; the latter contains besides, however, the field due to the dipole in question itself. The difference between $\boldsymbol{E}'$ and $\boldsymbol{E}$ would be vanishingly small only if a sufficiently small volume v with the electric moment $\boldsymbol{M}\,.\,v$ were to produce no appreciable field in its interior. But this is in no way true. Rather, the intensity of field which an arbitrary small sphere, polarized uniformly with the moment $\boldsymbol{M}$ per unit of volume, produces in its interior is everywhere in that interior, according to III (189), equal to $-\frac{4\pi}{3}\boldsymbol{M}$, that is, independent of the size of the sphere. It is only when we add this amount to the exciting field-strength $\boldsymbol{E}'$ that we obtain the total intensity of field $\boldsymbol{E}$, thus:

$$\boldsymbol{E}' - \frac{4\pi}{3}\boldsymbol{M} = \boldsymbol{E} \quad . \quad . \quad . \quad . \quad . \qquad (351)$$

If we combine the last three equations we get the relationship between the electric moment $\boldsymbol{M}$ and the

electric intensity of field $\boldsymbol{E}$, namely, the equation of vibration, in the form :

$$m\ddot{\boldsymbol{M}} + \left(m\omega_0^2 - \frac{4\pi}{3} Ne^2\right)\boldsymbol{M} = Ne^2\boldsymbol{E} \quad . \quad . \quad (352)$$

and, if we combine this with (348), we arrive at the desired generalization of the fundamental equation (4).

The expression on the left-hand side of the equation of vibration is often supplemented by a damping term in $\dot{\boldsymbol{M}}$ and a positive coefficient, as in I (20) or in III (375), which allows us to take into consideration any damping that may be caused either by the loss of electromagnetic energy or by collisions with neighbouring oscillators; we shall not, however, consider this extension here.

If we use the abbreviations :

$$\frac{4\pi Ne^2}{m} = a \quad . \quad . \quad . \quad . \quad . \quad . \quad (353)$$

$$\omega_0^2 - \frac{a}{3} = \omega'^2_0 \quad . \quad . \quad . \quad . \quad . \quad (354)$$

the equation of vibration assumes the simpler form :

$$\ddot{\boldsymbol{M}} + \omega'^2_0\boldsymbol{M} = \frac{a}{4\pi} \cdot \boldsymbol{E} \quad . \quad . \quad . \quad . \quad . \quad (355)$$

The equations (1), (2), (3), (348) and (355) form the basis of the theory of dispersion which we have described. This theory is largely due to H. A. Lorentz. Our first deduction from it is that $\omega_0^2 > \frac{a}{3}$, which imposes a certain upper limit on the density of distribution of a definite kind of oscillator. For otherwise $\boldsymbol{E}$ and $\boldsymbol{M}$ would have different signs in the statical field.

§ 79. In order first to link up with the original theory of Maxwell we consider processes which occur so slowly that in the equation of vibration (355) the term in $\ddot{\boldsymbol{M}}$ may be neglected in comparison with $\boldsymbol{M}$; this amounts to the frequency of the vibrations of the dipoles, caused

by the exciting wave, being small compared with their proper frequency. It then follows from (355) that:

$$\boldsymbol{M} = \frac{a}{4\pi\omega_0'^2}\boldsymbol{E}$$

and, if (348) is taken into account:

$$\boldsymbol{D} = \left(1 + \frac{a}{\omega_0'^2}\right)\cdot\boldsymbol{E}$$

Here, therefore, as in (4), we have strict proportionality between the electric induction and electric intensity of field, the factor of proportionality, namely, the dielectric constant, having the value:

$$\epsilon = 1 + \frac{a}{\omega_0'^2} \quad . \quad . \quad . \quad . \quad . \quad (356)$$

This is the region to which the consequences deduced in § 9 refer. The dispersion vanishes entirely and, by (22), we get for the refractive index:

$$n^2 = 1 + \frac{a}{\omega_0'^2} \quad . \quad . \quad . \quad . \quad . \quad (357)$$

The formulæ (356) and (357), which, as we see, contain a physical explanation of the nature of the dielectric constants, and hence also of the electric induction, may be tested experimentally by measuring the dependence of the dielectric constants or of the refractive index on the density of distribution N of the dipoles. We must note, however, that, by (353) and (354), a as well as $\omega_0'^2$ depends on N.

If we substitute the values in question in (357), we may write:

$$\frac{n^2 - 1}{n^2 + 2} = N \cdot \frac{4\pi e^2}{3m\omega_0^2} \quad . \quad . \quad . \quad . \quad (358)$$

Since the constants e, m and ω_0 are independent of N, this relationship is very convenient for determining the connection between the quantities n and N. Independently of H. A. Lorentz it has also been derived by the

Danish physicist L. Lorenz and is satisfactorily confirmed by measurement.

§ 80. For treating the general case it is advantageous to eliminate the magnetic intensity of field $\boldsymbol{H}$ and the electric induction $\boldsymbol{D}$ from the equations (1) and (348). In view of (2) this gives

$$\ddot{\boldsymbol{E}} + 4\pi\ddot{\boldsymbol{M}} = c^2\Delta\boldsymbol{E} \quad . \quad . \quad . \quad . \quad (359)$$

This equation together with (355) may serve for obtaining solutions of the dispersion problem.

CHAPTER II

PLANE WAVES

§ 81. PASSING on now to develop the most important laws of the electromagnetic processes that occur in a dispersive body from the fundamental equations that have been established, we follow along the line of reasoning adopted in § 3 in treating isotropic bodies and first consider a plane wave; that is, we assume that, of the three space co-ordinates x, y, z, only the co-ordinate x comes into action. We also assume the wave to be linearly polarized, the vibrations being in the y-direction, as formerly for the f-wave. Then $\boldsymbol{E}_x = 0$, $\boldsymbol{E}_z = 0$, $\boldsymbol{H}_x = 0$, $\boldsymbol{H}_y = 0$, and the equations (355) and (359) reduce to:

$$\ddot{\boldsymbol{M}}_y + \omega_0'^2 \boldsymbol{M}_y = \frac{a}{4\pi}\boldsymbol{E}_y \quad . \quad . \quad . \quad . \quad (360)$$

and:

$$\ddot{\boldsymbol{E}}_y + 4\pi\ddot{\boldsymbol{M}}_y = c^2\frac{\partial^2 \boldsymbol{E}_y}{\partial x^2} \quad . \quad . \quad . \quad . \quad (361)$$

Compared with the differential equation (6) for a body free from dispersion this equation exhibits the fundamental difference that it has no longer to satisfy the differential equations without introducing a definite assumption regarding the form of the wave-function. Nothing remains but to set up a particular solution of the equations and then, by combining a sufficient number of particular solutions, also to write down the general solution, exactly as was done in dealing with the problems of total reflection in § 12.

As there, we also assume:

$$\boldsymbol{E}_y = e^{i\omega\left(t - \frac{nx}{c}\right)} \quad . \quad . \quad . \quad . \quad . \quad . \quad (362)$$

and, correspondingly:

$$M_y = \alpha e^{i\omega\left(t - \frac{nx}{c}\right)} \quad . \quad . \quad . \quad . \quad . \quad (363)$$

This denotes a *singly periodic* wave with the velocity of propagation:

$$\frac{c}{n} = q \quad . \quad . \quad . \quad . \quad . \quad . \quad (364)$$

Hence we also call the constant n the refractive index, as in the case of bodies free from dispersion. A peculiarity in this case, however, is, as we shall see, that n can also be imaginary. Then, as in the case of total reflection, the wave is spatially damped.

The assumed values of E_y and M_y, when substituted in equations (360) and (361), satisfy them in actual fact provided that the following relationships hold between the constants:

$$n^2 - 1 = 4\pi\alpha = \frac{a}{\omega'^2_0 - \omega^2} \quad . \quad . \quad . \quad (365)$$

Since the constants a and ω'_0 are determined by the constitution of the body, there is accordingly for any frequency ω of the wave a definite index of refraction n and a definite ratio α of the electric moment M to the field-strength E. It then follows from (1) that the magnetic field-strength is given by:

$$H_z = nE_y \quad . \quad . \quad . \quad . \quad . \quad (366)$$

§ 82. Let us next enquire into the dependence of the refractive index n on the frequency ω of the wave. By (365) n is either real or purely imaginary. In the former case we choose n positive; that is, we make the wave proceed in the positive direction of x. For very small values of ω we obtain the relationship (357), as is obvious. If ω increases, n also increases, and, in fact, to an unlimited extent if ω approaches indefinitely close to the value ω'_0 (cf. Fig. 24). After this value has been exceeded n becomes purely imaginary and equal to $-i\kappa$, where κ is real and positive, since otherwise E_y would become

infinite for $x = \infty$. This lasts until ω reaches the value $\sqrt{\omega'^2_0 + a}$. In this interval κ decreases continuously from ∞ to 0, as is indicated in Fig. 24 by the dotted line. From then onwards the refractive index again becomes real and increases from zero asymptotically to the value 1, which it attains for $\omega = \infty$. The result is that the whole spectral region from $\omega = 0$ to $\omega = \infty$ is divided into three different parts (Fig. 24) by an "absorption band" which stretches from $\omega = \omega'_0$ to $\omega = \sqrt{\omega_0'^2 + a}$. To the left and to the right of the absorption band there is no absorption, but the refractive

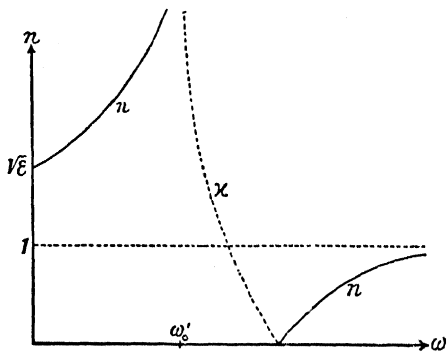

Fig. 24.

index always increases as the frequency increases; the dispersion is, as we say, "normal." But there is the essential difference that on the left, for smaller frequencies, the index of refraction is always greater than $\sqrt{\epsilon}$ (cf. (356)), whereas on the right, for greater frequencies, it is always less than 1. Within the absorption region there is no propagation of energy at all, the vibrations are stationary and are spatially damped, as in the case of total reflection in the optically less dense medium.

Concerning the ratio α of the amplitudes of the dipole-moment $\boldsymbol{M}$ and the field-strength $\boldsymbol{E}$, it is to be noted that by (365) it is positive on the left of the absorption band, but negative on the right and vanishes entirely when

$\omega = \infty$; that is, for infinitely rapid vibrations the dipoles play no part in the wave, the body behaves optically like a perfect vacuum.

§ 83. The laws of dispersion and absorption which have here been derived become a little more generalized if, in accordance with the idea already mentioned in § 78, a small damping term $k\dot{M}$, where $k > 0$, is added to the left-hand side of the equation of vibration (355). The term $k\dot{M}_{\nu}$ then becomes added in (360), and the same assumption (362) and (363), instead of leading to (365), leads to the values:

$$n^2 - 1 = 4\pi\alpha = \frac{a}{\omega'^2_0 - \omega^2 + ik\omega} \quad . \quad . \quad (367)$$

whereas the relationship (366) remains unaltered.

The only difference from the result previously obtained is that n is no longer either real or purely imaginary, but is always complex. If in place of n we now write generally $n - i\kappa$, then by (362) n again represents the real refractive index, κ the absorption index; and Fig. 24 is to be changed in that the curve for n no longer extends to infinity, then falling to zero and remaining zero in the whole region of absorption, but rather n remains continuous everywhere, rising to a steep maximum after entering the absorption region and then falling continuously to zero, corresponding to so-called "anomalous" dispersion, whereas conversely the curve for the absorption index κ keeps close to the axis of abscissæ in the two regions of normal dispersion and has appreciable values only in the region of the absorption band.

§ 84. A further generalization of practical importance enters into formula (367) if the dispersing body contains different kinds of simultaneously vibrating dipoles whose number per unit volume we shall denote by N_1, N_2, N_3 . . . In place of (348) we then have the relationship:

$$\frac{D - E}{4\pi} = \Sigma M_1 . \quad . \quad . \quad . \quad . \quad (368)$$

which must be summed up over all the kinds of dipoles; for each individual kind there is a vibration equation of the form (349), with or without a damping term, with constant coefficients ω_1, a_1, and possibly also k_1, while the equations (351) and (359) become generalized to:

$$\boldsymbol{E}' - \frac{4\pi}{3}\Sigma \boldsymbol{M}_1 = \boldsymbol{E} \quad . \quad . \quad . \quad . \quad (369)$$

$$\ddot{\boldsymbol{E}} + 4\pi\Sigma\ddot{\boldsymbol{M}}_1 = c^2\Delta\boldsymbol{E} \quad . \quad . \quad . \quad (370)$$

The same assumption (362) for the electric intensity of field and (363) for the dipole-moments $\boldsymbol{M}_1$, $\boldsymbol{M}_2$, . . . with the amplitudes α_1, α_2, α_3 . . . then leads, as in (367), to the general relationships:

$$\frac{n^2 - 1}{n^2 + 2} = \frac{1}{3}\sum \frac{a_1}{\omega_1{}^2 - \omega^2 + ik_1\omega} \quad . \quad . \quad (371)$$

and:

$$4\pi\alpha_1 = \frac{(n^2 + 2)a_1}{3(\omega_1{}^2 - \omega^2 + ik_1\omega)} \quad . \quad . \quad . \quad (372)$$

which determine the complex index of refraction $n - i\kappa$ and the amplitudes of the vibrating dipoles of all kinds for every wave-frequency ω. In geometrical language, as shown diagrammatically in Fig. 24, this relationship states that every kind of oscillator in the spectral region gives rise to a particular absorption band, caused by its proper constants, within which anomalous dispersion and appreciable absorption take place, whereas outside the absorption band the dispersion is always normal. But there is this difference, that in every intervening region between two absorption bands the real refractive index increases from very small to very great values, whereas in the outer region on the left (small values of ω) it is always greater than $\sqrt{\epsilon}$ and in the outer region on the right (great values of ω) it is always less than 1. In general we may say that every absorption-band whose frequency exceeds the wave-frequency ω, increases the real refractive index, whereas every absorption-band whose frequency is less than ω, reduces the refractive

index. Hence a refractive index which is greater than 1 always points to the presence of higher proper frequencies.

In all the following considerations we shall assume as known the complex refractive index $n - i\kappa$, by which the whole optical behaviour of the body is characterized.

§ 85. Concerning the laws of refraction and reflection of a plane periodic wave at the surface of a dispersive body or of two dispersive bodies in contact, these may be derived directly from the formulæ which hold for non-dispersive bodies. For if we reflect that the expressions which we have set up in (362) and (366) for the electric and for the magnetic field-strength of a wave which is progressing in the dispersive body are contained as special cases in the equations (8), and further, that the boundary conditions at the surface of separation, which express the continuity of the tangential field-components, are the same for dispersive as for non-dispersive bodies, it is clear that all the formulæ that were derived earlier for the reflected and the refracted wave also retain their validity here. The only difference is that here we must in general insert for the refractive index n and consequently, by (20), also for the angle of refraction θ_1 a complex value. This also makes the constants μ, σ, μ_1, σ_1, defined by the formulæ (23) and (24), of the reflected and refracted wave complex, as well as in the case of total reflection. The physical significance of this circumstance is expressed, as we saw earlier, by the occurrence of a sudden phase-change on reflection and refraction. Since this sudden change of phase is different for the wave which vibrates in the angle of incidence, the f-wave, from what it is in the case of the wave which vibrates perpendicularly to the plane of incidence, the g-wave, linearly polarized light is not linearly polarized after reflection, but elliptically polarized, and the measurement of the phase-difference of the two-wave components by means of a compensator, say a Fresnel rhomb (§ 27) or a sheet of mica (§ 64), leads to a relationship which, combined with the measurement of the refractive index,

gives us the two equations which are necessary if we wish to calculate the two constants n and κ, by which all the optical properties of the body are determined.

According to the general discussion of § 18 the intensity of radiation of the reflected and the refracted light is defined by the squares of the absolute values of the coefficients μ and σ, the quantities :

$$|\mu|^2 = \rho_{\parallel} \quad \text{and} \quad |\sigma|^2 = \rho_{\perp} \quad . \quad . \quad . \quad (373)$$

directly representing the reflection coefficients of the two wave-components, while in the case of the refracted wave the transmission coefficient is, by § 10, calculated more simply from the reflection coefficient than from the quantities μ_1 and σ_1. For the special case of normal incidence ($\theta = 0$) we get from (23) :

$$\rho_{\parallel} = \rho_{\perp} = \left|\frac{\theta - \theta_1}{\theta + \theta_1}\right|^2$$

$$= \left|\frac{(n_1 - i\kappa_1) - (n - i\kappa)}{(n_1 - i\kappa_1) + (n - i\kappa)}\right|^2 = \frac{(n_1 - n)^2 + (\kappa_1 - \kappa)^2}{(n_1 + n)^2 + (\kappa_1 + \kappa)^2} \quad (374)$$

From this we see among other things that the reflection coefficient vanishes only if both $n_1 = n$ and also $\kappa_1 = \kappa$, that is, if there is no optical difference at all between the two bodies in contact. Hence a " black body," that is, one which absorbs all light that falls on it, cannot have a plane surface. On the other hand, the reflection coefficient can very well approximate closely to 1, namely always when any one of the four quantities n_1, n, κ_1, κ is very great compared with the rest. In this case we speak of " metallic " reflection as in the case of an absolute conductor of electricity (III, § 92).

§ 86. The method which we have here adopted to derive the laws of dispersion and absorption from the hypotheses assumed concerning the nature and the disposition of the vibrating dipoles may be replaced by a totally different method, whose results may therefore simultaneously be used to test those just obtained. For instead of first setting up differential equations for

the electromagnetic phenomena that occur in a body by passing on from the assumption of an atomic constitution to that of a continuous constitution of the dispersive body and *then* forming particular integrals of these equations, there is essentially nothing to prevent our performing this transition later (indeed, we shall see that this deepens our comprehension of the phenomena in question) and first considering directly the electromagnetic waves that are emitted by the individual vibrating dipoles. On this view there is only one medium in which the waves propagate themselves, namely in the vacuum; in this vacuum there are embedded a great number of vibrating dipoles, each of which forms the centre of a spherical wave which propagates itself according to known laws independently of the remaining dipoles. For by III, 396 the external electromagnetic field produced by a dipole vibrating with the moment $f(t)$ is uniquely determined for all points of space and for all times. So we arrive at a particular solution of the problem, corresponding to the equations (362) and (363), if we assume the vacuum to contain firstly a periodic wave advancing with the velocity c in the direction of the x-axis and vibrating in the direction of the z-axis with the frequency ω—this is the "primary wave"—and secondly a great number of spherical waves, "secondary" waves, which are emitted by the individual dipoles in accordance with the vibrations which they emit with the same frequency; these secondary waves superpose themselves on each other and also on the primary wave. The result is the "effective" wave, whose field-strengths $\boldsymbol{E}$ and $\boldsymbol{H}$ are formed directly by vectorial addition of the corresponding field-strengths of the primary and the secondary waves. Since the dipoles lie very closely together, we may replace the addition by an integration.

Account is taken of the interactions of the dipoles among themselves and with the primary wave in that the equation (349) now holds, just as before, between the moment of vibration of a dipole and of the field-strength

$\boldsymbol{E}'$ which excites it, the equation (351) or (369) again holding for the relationship between $\boldsymbol{E}$ and $\boldsymbol{E}'$.

It is clear that these relationships give rise to a perfectly definite law for the propagation of an effective wave in the body in question; and if there is no inherent contradiction in any part of the theory, this law must be identical with that deduced earlier, although here we have made use neither of the differential equations of dispersion nor of the conception of electric induction, but only of the laws of waves in a pure vacuum.

This postulate is fulfilled in actual fact, as we can easily show by direct calculation. It must be particularly emphasized that compounding the secondary spherical waves due to the dipoles vectorially must give and does give two results—firstly a wave which advances with the velocity c and is equal and opposite to the primary wave, which also advances with the velocity c and hence just neutralizes it; secondly a wave which advances with the velocity $\frac{c}{n}$, where n is determined by (365) or (371) respectively. In all directions other than that of the x-axis the secondary waves destroy each other by interference.

We see that this new method of regarding the problem gives us a direct understanding of the nature of dispersion. But its physical significance goes still further. For it also leads to a new derivation of the laws of refraction and reflection. Let us imagine, for simplicity, a plane periodic wave to fall from a vacuum on to the plane surface of a dispersive body at the angle of incidence θ. This wave, which we may again call the primary wave, passes completely undisturbed through the body; for it propagates itself with the velocity c in the body just as in the vacuum. According to this view, then, there is no meaning in speaking of a surface of separation or contact. The influence of the body makes itself felt solely by the simultaneous vibration of its dipoles, and we may again regard the effective wave-process as the

superposition of the primary wave on the secondary waves emitted by the vibrating dipoles of the body, by making the appropriate assumptions for the moments of the dipoles, which must of course satisfy the equation of vibration (349). The result of the calculation again confirms in every detail the laws derived in the preceding section by a totally different method. For the result of superposing the secondary spherical waves is firstly a plane wave which completely destroys the primary wave by interference; and secondly a plane wave which advances in the direction of the ray refracted according to Snell's law, and thirdly a wave which is reflected back into the vacuum at the angle θ. The amplitudes of the waves also come out as above calculated.

Although this method of deriving the laws of reflection and refraction is more complicated mathematically, it is more significant physically than the former method. For since it involves no boundary condition it gives a direct and complete explanation of the phenomena of reflection and refraction, which, according to it, are nothing but the combined effects of spherical waves that have been emitted and interfere in a particular way.

§ 87. All the preceding discussion refers exclusively to such particular solutions of the field-equations as correspond to singly periodic waves. If we now turn more generally to the investigation of *non-periodic* plane waves which advance in the x-direction within a dispersive body, it must be emphasized that the conception of velocity of propagation here loses completely the general meaning which it has for non-dispersive bodies. For in a dispersive body it is not always possible to represent the wave-function as a definite function of a single argument which depends only on x and t, as was done in equations (8); or, in other words, the wave does not advance unchanged, but becomes deformed in the course of time, and it is impossible to recognize with certainty any selected point of the wave again at a later time.

From this it follows at once that we are at liberty to

define an infinite number of different kinds of velocities of propagation according to the characteristic feature that we fix on as the wave advances. For example, we can fix on the value of the field-strength $\boldsymbol{E}_y$ at a point x at any time t and enquire what is the neighbouring point $x + dx$ at which the field-strength has the same value at the time $t + dt$ ($\boldsymbol{E}_y$ = const.). This gives the following value for the velocity of propagation of the field-intensity $\boldsymbol{E}_y$:

$$\frac{dx}{dt} = -\frac{\dfrac{\partial \boldsymbol{E}_y}{\partial t}}{\dfrac{\partial \boldsymbol{E}_y}{\partial x}} \quad . \quad . \quad . \quad . \quad . \quad (375)$$

a value which is always constant for non-dispersive mediums, but for dispersive mediums it is constant only for singly periodic waves, and hence in general has no deeper significance. The same applies if we consider the velocity of propagation of a maximum or a minimum of the field-strength. Its value is obtained from (375) if we substitute $\frac{\partial \boldsymbol{E}_y}{\partial x}$ in it for $\boldsymbol{E}_y$ and is of course different in general from (375).

It follows from this discussion that it is altogether impossible to define for a dispersive body a velocity of propagation which is constant in space and in time, so long as we make no assumption about the special form of the wave. It is easy to see, moreover, that for different forms of the wave the velocity of propagation may be defined in quite different ways. We shall see this confirmed in several particularly important cases.

If we now enquire into the general integral of the wave equation (360) and the associated equation of vibration (361) we can find an essentially adequate solution by the same method as was successful in the case of the problem of total reflection, in § 14. For we may express any arbitrary function over any arbitrarily great region of its argument as a Fourier series, and then every indi-

vidual term of the Fourier series represents a singly periodic occurrence which obeys the laws of propagation found above. Since the differential equations (360) and (361) are linear and homogeneous, the individual processes become superposed on each other and hence it follows that every wave which advances in a dispersive body may be regarded as composed of a number of singly periodic waves each of which advances unchanged with the velocity of propagation $\frac{c}{n}$ which is characteristic of it and which is constant in space and in time. But since n is different for the individual waves, the phases of the individual partial wave become displaced relatively to one another and so cause the deformation of the total wave.

§ 88. We shall now fix our attention on the special case where the dispersive body, which we shall suppose stretches from $x = 0$ to $x = \infty$, is perfectly unexcited at first and that from a certain moment of time onwards, say $t = 0$, a singly periodic wave of frequency ω is continuously incident on its surface in the normal direction. With what velocity will the disturbance propagate itself into the body? We may be tempted to assume at first sight that the wave simply propagates itself into the body with the velocity $\frac{c}{n}$, as is the case with a permanently periodic wave. But this is not so. For here we are concerned with a non-periodic wave which becomes deformed as it advances. The complete solution of this problem has been given by Sommerfeld,[1] who resolved the waves into a Fourier integral. But we can find the answer to the above question without making special calculations if we apply the line of reasoning which was fully discussed in § 86.

The primary wave also advances in the dispersive body with the velocity c. The secondary waves due

[1] A. Sommerfeld, Weber-Festschrift, B. G. Teubner, p. 338, 1912.

to the vibrating dipoles do the same. But the result of this joint action is not the same as in the case of a permanently periodic wave. For the dipoles do not vibrate periodically; they were at rest initially and, on account of their inertia, they will only gradually be made to vibrate, the later the greater their abscissæ x, and these vibrations, strictly speaking, will acquire their constant character only after an infinitely long period of time and so give rise to the singly periodic wave-motion whose velocity of propagation is $\frac{c}{n}$. So long as this is not the case the primary wave will not be neutralized by the secondary waves and so it manifests itself as a disturbance which advances in the dispersive body with the velocity c. Since on the other hand the secondary waves, no matter what their form may be, also propagate themselves with the velocity c it follows that the desired velocity with which the head of the wave advances into the body, the "front velocity," is always equal to c. In contrast with the front velocity is the "phase-velocity," $\frac{c}{n}$, which a permanently singly periodic wave has; in future we shall denote it by u:

$$\frac{c}{n} = u \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (376)$$

We must not regard it as an inherent contradiction to the theory that the phase-velocity u may also be greater than the front velocity c $(n < 1)$. For the wave is not singly periodic and the head of the wave that penetrates into the body has no constant phase. Hence there can be no question of the head of the wave which advances with the front velocity being passed by the hinder part of the wave which advances with the phase-velocity.

It further follows from this that the velocity of propagation of a light-signal emitted into the body, the "signal velocity," can at most be equal to c. This maximum value is attained in an ideal detector, that is, a receiver which reacts to even the smallest disturbance. Other-

wise the signal velocity must be made correspondingly smaller according to the sensitivity of the receiver. The value of the phase-velocity does not come into question at all in the matter; for it is not possible to send signals by means of singly periodic waves.

§ 89. We shall now fix our attention on a non-periodic wave of a form which is in principle still more important, namely a wave which is *nearly singly periodic*. For this case is realized in nature in all so-called monochromatic rays which are actually produced (§ 16). The general expression for a wave of this kind which is nearly periodic and which advances in a dispersive body is, by (362):

$$E_y = \sum_{m=1}^{N} C_m e^{i\omega_m\left(t - \frac{n_m x}{c}\right)} \quad . \quad . \quad . \quad (377)$$

where:

$$\frac{\omega_N - \omega_1}{\omega_1} \ll 1 \quad . \quad . \quad . \quad . \quad (378)$$

Here m is the order number with which the frequency ω_m increases, the total number N is arbitrarily great and C_m is any complex constant. Let us first consider the form of such a wave, that is, E_y as a function of x at a definite time t, say represented by a curve with the abscissa x. This curve is formed by the superposition of a number of simple sine curves of very nearly the same wave-length. The result is, as emerged from our exhaustive discussion in § 16, a nearly periodic curve of the same wave-length, but with a slowly and in general irregularly changing amplitude and phase.

Let us consider the conditions a little more closely. The decisive features for the amplitude and phase of the resultant wave are the phase-differences with which the individual singly periodic partial waves meet. If we fix on any one of them of frequency ω and phase-velocity $\frac{c}{n}$, then the phase of any other partial wave of frequency ω' and phase-velocity $\frac{c}{n'}$, can, by (377), for all

points x and all times t, be set equal to the phase of the first wave plus the phase-difference :

$$(\omega' - \omega)t - \frac{x}{c}(\omega' n' - \omega n) + \theta' - \theta \quad . \quad . \quad (379)$$

where θ' and θ denote the angular arguments of the complex constants C' and C. So long as this phase-difference remains appreciably unchanged for all combinations of any two partial vibrations, the partial waves interfere in the same way and the resultant wave has a constant amplitude and phase. Now if the time t is assumed constant the phase-difference changes only with the abscissa x. If we pass from the point x to a neighbouring point $x_1 > x$ the phase-difference (379) changes by :

$$-\frac{x_1 - x}{c}(\omega' n' - \omega n) \quad . \quad . \quad . \quad (380)$$

Hence so long as the two points lie so close together that for all pairs of partial vibrations :

$$x_1 - x << \left| \frac{c}{\omega' n' - \omega n} \right| \quad . \quad . \quad . \quad (381)$$

the resultant wave forms between the two points a simple sine curve of wave-length :

$$\lambda = \frac{2\pi c}{n\omega} \text{ or, respectively, } \frac{2\pi c}{n'\omega'} \quad . \quad . \quad (382)$$

which, by (378), are appreciably the same. But if we pass so far along the axis of abscissæ with the point x_1 that the relationship (381) ceases to be valid, the interference effect of the partial waves changes and the resulting wave acquires a new amplitude and a new phase-constant. Accordingly we may say : the form of the resultant wave $\boldsymbol{E}_y$ at any time t may be regarded in general as compounded of a number of successive groups of simple sine-curves. Every single group stretches over a region of abscissæ for which the relationship (381) is fulfilled ; a characteristic of the group is a definite phase-

difference (379) for any two partial vibrations, and consequently also a definite amplitude and a definite phase-constant. All these quantities vary from group to group. The number of wave-periods that belong to a group is, by (382):

$$\frac{x_1 - x}{\lambda} = \frac{(x_1 - x) \cdot n\omega}{2\pi c} \quad . \quad . \quad . \quad (383)$$

where x and x_1 denote the initial and the final points of the group. This number may be very large; for the relationship:

$$\frac{(x_1 - x) \cdot n\omega}{2\pi c} >> 1 \quad . \quad . \quad . \quad . \quad (384)$$

is, on account of (378), certainly compatible with the relationship (381).

Let us enquire further into the changes that occur in the course of the time t firstly when we keep a definite point x in space fixed. Then, by reasoning which is fully analogous to that given above, we find that within an interval of time $t_1 - t$ which is so small that for every pair of partial vibrations:

$$t_1 - t << \left| \frac{1}{\omega' - \omega} \right| \quad . \quad . \quad . \quad . \quad (385)$$

the resultant field-strength E_y executes pure sine vibrations of frequency ω, but that at later times t a change in the amplitude and the phase-constant occurs.

The process is a little easier to picture if we fix our attention not on a definite point x of space but on a definite phase of the resultant wave, say the peak of the crest of a wave. This crest advances with the phase-velocity $\frac{c}{n}$ and retains its value and its velocity of propagation along a certain distance which may include many wave-lengths. But after that it will slightly change its value and also its velocity of propagation. The latter will always be nearly equal to the phase-velocity, but the value of the crest of the wave can increase or decrease

by an arbitrary amount. In other words, any definite wave-crest on which we fix our attention does not remain in the same group in the course of time, but passes over into a neighbouring group.

This circumstance leads us to enquire with what velocity a definite selected group advances. The answer is contained in the above given property of a group that for it the phase-difference (379) for every combination of any two partial vibrations retains a definite invariable value. Hence if we set the value (379) for x, t equal to that for x_1, t_1, we get:

$$\frac{x_1 - x}{t_1 - t} = \frac{(\omega' - \omega)c}{\omega' n' - \omega n} \quad . \quad . \quad . \quad . \quad (386)$$

which is independent of x and t. This velocity, which is constant for all points and all time and which may, on account of (378), also be written in the form:

$$c \cdot \frac{d\omega}{d(\omega n)} = v \quad . \quad . \quad . \quad . \quad . \quad . \quad (387)$$

represents the required group-velocity. It is connected with the phase-velocity u, according to (376), by the relationship:

$$\frac{1}{v} = \frac{d\left(\frac{\omega}{u}\right)}{d\omega} \quad . \quad . \quad . \quad . \quad . \quad . \quad (388)$$

The difference in u and v of course depends on the circumstance already pointed out above, that a definite selected wave-crest advancing with its phase-velocity changes from group to group in the course of time. It is only in the case of dispersionless bodies $\left(\frac{dn}{d\omega} = 0\right)$ that $v = u$, and every wave-crest remains permanently in its group. In the case of normal dispersion $\left(\frac{dn}{d\omega} > 0\right)$ we have $v < u$, that is, in the course of time a definite wave-crest moves out of its group forwards into the next group and so catches up one group after another. In

the case of anomalous dispersion the reverse occurs. But we must bear in mind that a group does not form a configuration which is permanent for all time. For since the differential-quotient in (386) is not, strictly speaking, equal to the differential quotient in (387) but has a different value for every combination of two partial vibrations, in a fairly great interval of time $t_1 - t$ the distance $x_1 - x$ traversed will be appreciably different for different combinations, that is, the group will fall apart. Hence in rather great intervals of time a gradual change also occurs between the groups in that every group eventually splits up and fuses partly with the preceding group and partly with the following group.

CHAPTER III

GEOMETRICAL OPTICS OF NON-HOMOGENEOUS BODIES. RELATIONSHIPS TO QUANTUM MECHANICS

§ 90. An optically non-homogeneous body is characterized in having its refractive index n dependent in some definite way on the space-co-ordinates x, y, z. We shall also assume that the body is dispersive, that is, we also make n depend on the frequency ω. But we shall assume that it is isotropic, so that at a definite point and for a definite frequency n is the same in all directions. The laws of propagation of light-waves for such a body are in general very complicated, but an essential simplification may be achieved by restricting ourselves to the region of geometrical optics or ray-optics which, by § 28, embraces those phenomena for whose representation the laws which hold for plane waves suffice.

Let us again begin with singly periodic vibrations. Then every wave-function has the phase:

$$\omega t - \phi(x, y, z) = \Theta \quad . \quad . \quad . \quad . \quad (389)$$

where ω denotes the frequency, and ϕ a certain function of position, the "eikonal"; the conception of ϕ is generalized from (362).

If we define a wave-surface as a surface whose points at a definite time t are all in the same phase Θ, then a wave-surface is represented by the equation:

$$\phi(x, y, z) = \text{const.} \quad . \quad . \quad . \quad . \quad (390)$$

The family of wave-surfaces which results when we allow the constant to run through all its values is the same for all times and hence is fixed in the body. If we fix on a

definite phase Θ, then, by (389), it will be found at the time t in the surface $\phi = \omega t - \Theta$, and so as the time increases it passes uniformly from smaller to greater values of ϕ. If, on the other hand, we fix on a definite point x, y, z in space, the phase at it increases uniformly with the time:

$$\frac{\partial \Theta}{\partial t} = \omega \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (391)$$

The characteristic theorem of geometrical optics is that every sufficiently small portion of a wave-surface can be regarded as a plane wave which propagates itself independently in the direction of its normal with the phase-velocity $u = \frac{c}{n}$ which it has at the point in question. Since the body is isotropic, the direction of the normal is simultaneously that of the ray; and, if any particular wave-surface is known, we obtain an infinitely near wave-surface by marking off from every point of the given wave-surface the distance $u \, . \, dt$ in the direction of the normal and then connect the end-points of all these distances together. Thus if ds is the normal distance between the two surfaces, we have:

$$ds = udt.$$

Since, on the other hand, the phase Θ remains the same during the propagation of the wave—for u is the phase-velocity—we have, by (389):

$$\omega dt - d\phi = 0$$

and this, combined with the preceding equation, gives the following value for the differential coefficient of ϕ in the direction of the normal:

$$\frac{\partial \phi}{\partial s} = \frac{\omega}{u} = \frac{n\omega}{c} = -\frac{\partial \Theta}{\partial s} \quad . \quad . \quad . \quad . \quad (392)$$

By means of this differential equation, the so-called eikonal equation, it is possible to calculate from a single wave-surface all the remaining wave-surfaces, since n is

given as a function of x, y, z. The whole of geometric optics is in essence contained in this equation. For if the system of wave-surfaces is known, its orthogonal curves represent the rays which propagate themselves through the body.

§ 91. The laws of geometric optics, besides allowing themselves to be expressed by equation (392), may also be represented by Fermat's principle, the derivation of which may be made clear in a way similar to that used in § 37 for a special case.

According to Fermat's principle, the time which a definite phase requires to arrive with its characteristic velocity u from a definite point P to a definite other point Q of the body along the path of the ray which goes from P to Q is less than that taken along any other line connecting P and Q, or, symbolically:

$$\delta \int_P^Q \frac{ds}{u} = 0 = \delta \int_P^Q n ds \quad . \quad . \quad . \quad . \quad (393)$$

If we substitute for the variations in the equation:

$$\delta n = \frac{\partial n}{\partial x}\delta x + \frac{\partial n}{\partial y}\delta y + \frac{\partial n}{\partial z}\delta z$$

$$\delta ds = \frac{dx}{ds}\delta dx + \frac{dy}{ds}\delta dy + \frac{dz}{ds}\delta dz$$

and then, as in I, § 108, integrate by parts so as to reduce the variations δdx, δdy, δdz to δx, δy, δz, afterwards setting the coefficients individually equal to zero, the differential equations for the path of the ray from P to Q is obtained in the form:

$$d\left(n\frac{dx}{ds}\right) : d\left(n\frac{dy}{ds}\right) : d\left(n\frac{dz}{ds}\right) = \frac{\partial n}{\partial x} : \frac{\partial n}{\partial y} : \frac{\partial n}{dz} \cdot \quad . \quad (394)$$

These equations, by which the path of the rays can be calculated if the starting-point and the direction of the ray are given, also contain all the laws of geometric optics. In content they are completely identical with equation (392) for the wave-surfaces, as can also be seen

directly if we imagine the construction described in § 90 to be carried out for a wave-surface adjacent to a given wave-surface and if we calculate the resultant change in the direction of the normal, that is, the difference between the normals of the new and the original surface. We then arrive at exactly the relationship (394).

§ 92. In order to be sure of our subsequent procedure we shall first formulate a little more accurately the conditions under which geometric optics may be applied. In general the differential equations of dispersion (359) and (355) hold. The former contains only the universal constant c, in the latter we assume for a non-homogeneous body that the constants ω_0' and a are given functions of the space-co-ordinates x, y, z.

The two differential equations are then satisfied by a generalization of the assumptions (362) and (363) made for plane waves in a homogeneous body :

$$E = \psi, \quad M = \alpha\psi \quad . \quad . \quad . \quad . \quad (395)$$

where α is now to be a certain function of position, while ψ, the wave-function, depends in some way on phase and time.

By substituting in (355) and (359) we obtain :

$$\alpha\ddot{\psi} + \omega_0'^2\alpha\psi = \frac{a}{4\pi}\psi \quad . \quad . \quad . \quad . \quad (396)$$

and:

$$\ddot{\psi} + 4\pi\alpha\ddot{\psi} = c^2\Delta\psi$$

or:

$$\ddot{\psi} = \frac{c^2}{n^2}\cdot\Delta\psi = u^2\Delta\psi \quad . \quad . \quad . \quad . \quad (397)$$

where n is again related to α by (365).

If we now introduce the assumption that the process is singly periodic in time, that is :

$$\psi = e^{i(\omega t - \phi)} \quad . \quad . \quad . \quad . \quad . \quad (398)$$

where the function of position ϕ, being a generalization from (389), may also be complex, then :

$$\ddot{\psi} = -\omega^2\psi \quad . \quad . \quad . \quad . \quad . \quad (398a)$$

and the relationship (396) leads to the same value of n as the relationship (365) deduced for plane waves in homogeneous bodies. So the whole problem is in general reduced to a differential equation (397) in the wave-function ψ.

Substituting the equation (398) we get:

$$\frac{n^2\omega^2}{c^2} = i\Delta\phi + \left(\frac{\partial\phi}{\partial x}\right)^2 + \left(\frac{\partial\phi}{\partial y}\right)^2 + \left(\frac{\partial\phi}{\partial z}\right)^2$$

or, if ds again denotes the element of length of the normal to the surface ψ = const.:

$$\frac{n^2\omega^2}{c^2} = i\Delta\phi + \left(\frac{\partial\phi}{\partial s}\right)^2 \quad . \quad . \quad . \quad . \quad (399)$$

This relationship becomes the same as the fundamental equation (392) of geometric optics if and only if the term in $\Delta\phi$ can be neglected. Since the equation (399) is complex, two conditions follow from this. Firstly the differential coefficients of the imaginary part of ϕ must be very small or, what comes to the same thing, according to (398), the amplitude of the wave-function must be only slowly variable in space; secondly, the term in the second differential coefficient of the real part of ϕ must be small compared with each of the other two terms of the equation (399). In view of (382) this means that the radii of curvature of the wave-surfaces must be great compared with the wave-length—the same postulate as we set up earlier in § 28. Or, the radii of curvature of the rays must be great compared with the wave-length. So long as these conditions are fulfilled, the laws of geometric optics may be applied as a first approximation. But as soon as they are transgressed at any point the exact wave-equation (397) enters into force in place of them.

§ 93. We now turn from singly periodic rays to the natural monochromatic rays, that is, nearly singly periodic rays of the same kind as those we investigated in § 86. We saw there that such a ray may be regarded as a

progressive series of consecutive groups of singly periodic rays whose velocity of propagation v is characterized by the condition that the difference $\Theta' - \Theta$ of the phases of any two partial vibrations ω' and ω remains constant. Thus by (389):

$$(\omega' - \omega) \cdot t - (\phi' - \phi) = \text{const.}$$

or:

$$(\omega' - \omega) \cdot dt - \left(\frac{\partial \phi'}{\partial s} - \frac{\partial \phi}{\partial s}\right) ds = 0.$$

From this we get, taking (392) into account:

$$\frac{ds}{dt} = v = \frac{\omega' - \omega}{n'\omega' - n\omega} c$$

which agrees perfectly with (386), (387) and (388).

Since the group velocity v is different from the phase-velocity u, a point which advances with the group-velocity will in the course of time change its phase, and by (389) we have:

$$\frac{d\Theta}{dt} = \frac{\partial \Theta}{\partial t} + \frac{\partial \Theta}{\partial s} v$$

and by (391) and (392):

$$\frac{d\Theta}{dt} = \omega - \frac{v\omega}{u} = \omega\left(1 - \frac{v}{u}\right) \quad . \quad . \quad . \quad (400)$$

Combined with (376) and (387) we get from this:

$$\frac{d\Theta}{dt} = \frac{\omega \dfrac{\partial n}{\partial \omega}}{\dfrac{\partial n}{\partial \omega} + \dfrac{n}{\omega}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (401)$$

The differential coefficients have here been written with round ∂'s because n also depends on position.

For a non-dispersive body $\frac{\partial n}{\partial \omega} = 0$, so that in this case the phase Θ remains constant, that is, the group-velocity

coincides with the phase-velocity. This holds for non-homogeneous as well as for homogeneous bodies.

§ 94. The fundamental equations of geometric optics above derived for a non-homogeneous dispersive isotropic body for the case of a nearly singly periodic ray exhibit in their formal structure a remarkable analogy with the fundamental equations of classical mechanics for a free point-mass which moves in a given statical conservative field of force (I, § 49)—an analogy which expresses itself in the circumstance that a wave-group (restricted to a small portion of space) moves in the direction of its ray according to the same laws as the point-mass moves in the direction of its orbital curve. It is, indeed, possible, by appropriately allocating the optical constants in the first case to the mechanical constants in the second case, to make the optical and the mechanical equations identical. This will now be shown.

For this purpose we first consider our two equations (391) and (401), which specify the time-change of phase Θ for a stationary point and for a point moving along the ray with the group-velocity v. We compare these equations with the two corresponding equations of mechanics for the integral of action W, considered as a function of x, y, z, t, namely (I, 420):

$$\frac{\partial W}{\partial t} = -E \quad . \quad . \quad . \quad . \quad . \quad (402)$$

for a stationary point, and (I, 419):

$$\frac{dW}{dt} = E - 2U \quad . \quad . \quad . \quad . \quad (403)$$

for a point which moves on its orbital curve according to the laws of motion. Here E is the energy and $E - 2U$ is the kinetic potential or Lagrange's function, namely, the difference of the kinetic energy:

$$E - U = \frac{1}{2}mv^2 \quad . \quad . \quad . \quad . \quad (404)$$

and the potential energy U, which we consider as a given function of the co-ordinates x, y, z.

In addition to the formulæ for the time derivatives of Θ on the one hand and W on the other we have those for the space derivatives, namely, equation (392) for Θ and the equations I (417) for W, which express that the space gradient of W is equal to the momentum of the moving point-mass, that is, that the direction of the velocity at every moment is perpendicular to the surface $W = \text{const.}$ and that the value of the gradient $\frac{\partial W}{\partial s}$ represents the product of the mass and the velocity :

$$\frac{\partial W}{\partial s} = mv \quad . \quad . \quad . \quad . \quad (405)$$

If we now identify the velocity v of the moving point-mass in size and direction with the velocity of propagation of the wave-group designated by v earlier, and if we reflect that the ray-direction is perpendicular to the wave-surface $\Theta = \text{const.}$, we shall be led to set the functions W and Θ proportional to one another. It then follows by comparison with (391) and (402) that the optical constant ω is proportional to the mechanical constant E. Since these two quantities have different dimensions, the factor of proportionality is not a pure number. Therefore we set :

$$E = h \cdot \frac{\omega}{2\pi} \quad . \quad . \quad . \quad . \quad . \quad . \quad (406)$$

and interpret h as a certain constant whose dimensions are the product of an energy and a time. Then (391) and (402) become identical, if we further set :

$$W = -\frac{h}{2\pi} \cdot \Theta \quad . \quad . \quad . \quad . \quad . \quad (407)$$

This also establishes the relationship between the equations (405) and (392) :

$$mv = \frac{\partial W}{\partial s} = -\frac{h}{2\pi}\frac{\partial \Theta}{\partial s} = \frac{n\omega h}{2\pi c} = \frac{nE}{c} \quad . \quad . \quad (408)$$

and from this we get, if v is calculated from (404), the following value of the refractive index :

$$n = \frac{c}{E} \cdot \sqrt{2m(E - U)} \quad . \quad . \quad . \quad (409)$$

The dependence of the refractive index on the function of position U is caused by the non-homogeneity of the body, the dependence on the energy constant E or, respectively, ω, is due to the dispersion of the body in which the wave-group in question is propagated.

It is now easy to convince ourselves of the fact that through the relationships which have been introduced between the mechanical and the optical quantities the two systems of equations become fully identical. In particular the equation (387) for the group-velocity :

$$\frac{1}{v} = \frac{1}{c}\frac{\partial(\omega n)}{\partial\omega} = \frac{1}{c}\frac{\partial(En)}{\partial E} \quad . \quad . \quad . \quad (410)$$

leads identically to the value of v in (404), if we substitute from (409), and in the same way (401) transforms directly into (403).

So we may perfectly generally enunciate the theorem: according to the laws of classical mechanics a point-mass m moves in a field of force of given potential energy U exactly like a singly periodic ray in a non-homogeneous dispersive isotropic body whose index of refraction is :

$$n = \frac{2\pi c}{\omega h}\sqrt{2m\left(\frac{\omega h}{2\pi} - U\right)} \quad . \quad . \quad . \quad (411)$$

Just as the motion of the point-mass is determined by the initial position, the initial direction and the energy E, so the propagation of the ray corresponding to it is determined by the initial point, the initial direction and the frequency $\omega = \dfrac{2\pi E}{h}$.

Although a far-reaching agreement is established in this way, between the laws of classical mechanics and those of geometrical optics, the agreement is not perfect.

For there is a fundamental difference in the following respect. Classical mechanics demands absolute validity: it is complete in itself and is not bound by assumptions concerning orders of magnitude. Of geometrical optics we know, however, that it does not hold generally, but by § 92 only if the wave-length (382) corresponding to the frequency ω of the ray is small compared with the radius of curvature of the path of the ray. As soon as this condition is infringed the laws of geometrical optics become inadequate and we have to fall back on the general wave-equation (397) or (399).

§ 95. We would be able to rest satisfied with these results if classical mechanics accounted for all the dynamical processes that occur in nature. But this is not the case.

Rather the facts of experience compel us to assume that in the case of very small quick movements certain of the circumstances arise which are quite foreign to the basic assumptions of classical mechanics; for example, in the motion of an electron about an atomic nucleus certain perfectly definite values of the energy E appear to be favoured—a theorem which is totally foreign to classical theory.

Hence if it has been shown to be necessary to change classical mechanics, it can only be a question of generalizing the laws hitherto established for it; for these laws would then retain their significance under all circumstances. Now we must bear in mind that classical mechanics can never be generalized from within itself; for, like every system that is perfect in itself, it bears the stamp of inner completeness and conceals no germs which allow of further development. Rather, the impulse must come from without. A beautiful example of this is given by the theory of relativity, which, starting from optics, has penetrated into mechanics and has added a totally new side to it.

In the present case, too, help has come from the direction of optics. L. de Broglie first adduced the long

familiar analogy, discussed in the previous section, between classical mechanics and geometric optics to account for quantum phenomena, and E. Schrödinger then developed the idea further, showing that the extension which the laws of classical mechanics must incur in order to adapt themselves to reality is of exactly the same type as that which must be applied to geometric optics when the limits within which it holds are overstepped. In other words, the laws of the new mechanics are found simply by retaining throughout the analogy of mechanics with optics and accordingly modifying the laws of the old mechanics so that the basic difference mentioned at the end of the previous section is eliminated. The equations (406) and (411), which are characteristic for the analogy between mechanics and optics, then remain valid beyond the range of validity of geometric optics, and out of classical mechanics, ray mechanics, we thus get wave mechanics, which contains the former as a special case. Its characteristic feature is that it no longer allows the motion of a point-mass to be represented by the motion of a geometrical point, but that the point-mass is in a certain sense resolved into a number of waves of a definite kind, just as a light-ray, strictly speaking, does not allow itself to be represented by a single curve, but arises only from the combined action of optical waves.

The first great success of this idea is shown by introducing the wave-equation (397) into the new mechanics. This gives with the value of (n) in (411):

$$\frac{8\pi^2 m}{\omega^2 h^2}\left(\frac{\omega h}{2\pi} - U\right)\ddot{\psi} = \Delta\psi \quad . \quad . \quad . \quad (412)$$

or, if we substitute the expression (398a) for $\ddot{\psi}$:

$$\Delta\psi + \frac{8\pi^2 m}{h^2}\left(\frac{\omega h}{2\pi} - U\right)\psi = 0 \, . \quad . \quad . \quad (413)$$

for which, by (406), we may also write:

$$\Delta\psi + \frac{8\pi^2 m}{h^2}(E - U)\psi = 0 \, . \quad . \quad . \quad (414)$$

The peculiarity of this differential equation of Schrödinger consists in the fact that the coefficient $(E - U)$ can also become negative in some circumstances. If the constant E is taken so great that for all points of the space—and the wave has no spatial limits—$(E - U) > 0$, there are finite and continuous solutions of the equation for all values of E. But if E is not sufficiently great to make the coefficient positive for even the greatest values of U, there are finite and continuous functions ψ everywhere, which satisfy the equation (414) only if E has certain definite values E_1, E_2, E_3, —, the so-called proper values (*Eigenwerte*). Corresponding to these values we have the proper functions (*Eigenfunktionen*) ψ_1, ψ_2, ψ_3 . . . Perhaps this startling result appears less striking if we consider that for negative values of $(E - U)$ the value (409) of the refractive index becomes imaginary and so geometric optics loses its meaning entirely.

But the fact that gives the differential equation (414) its great importance is that the proper values for E calculated from it agree exactly in all cases with those proper values which had already been calculated earlier by Heisenberg, Born and Jordan from the equations of quantum mechanics in matrix form; these equations had been based directly on experimental facts and had been developed independently of any particular physical assumptions. This subsequent agreement between results which had been obtained by two totally independent methods is a definite indication of their physical significance, and hence there can be no doubt that Schrödinger's differential equation, in virtue of its close relationship to classical mechanics, deepens our insight into the nature of quantum phenomena.

It is true that much still remains to be done before we shall be clear about the physical nature of the wave-function ψ and before we can satisfactorily answer all the questions that thrust themselves upon us.

§ 96. We shall conclude by showing how it is possible to calculate the numerical order of magnitude of the

universal constant h in (406) by using the hypothesis that was introduced in the preceding paragraph. In the first case it is clear that if we assume h to be infinitely small, wave-optics becomes merged into geometric optics, and hence quantum mechanics becomes identical with classical mechanics. For with this assumption it follows from (406) that for every value of E there is an infinitely great value of ω, and hence by (382) an infinitely small value for the wave-length λ; so the condition that the wave-length must be small compared with the radius of curvature of the orbital curve is always fulfilled.

From this we see that the value of h will come out the greater the greater the radii of curvature of the orbital curve at which the deviations from classical mechanics become appreciable. We shall now take as the basis of our estimate the assumption that for an electron which revolves around a stationary atomic nucleus which bears a single positive charge the deviation from classical mechanics becomes appreciable when its orbital radius shrinks to atomic size.

We have by classical mechanics in general that if r denotes the orbital radius, v the orbital velocity, e the charge, m the mass of the electron, then in the case of uniform circular motion the centripetal force is equal to the attractive force of the nucleus, thus :

$$\frac{mv^2}{r} = \frac{e^2}{r^2}$$

and :

$$v = \frac{e}{\sqrt{mr}} \qquad \qquad (415)$$

On the other hand, we get for the wave-length λ which is to be allocated to the motion here considered, by (382), (411), (406) and (404) :

$$\lambda = \frac{h}{\sqrt{2m(E-U)}} = \frac{h}{mv}$$

and by (415):

$$\lambda = \frac{h}{e}\sqrt{\frac{r}{m}} \quad . \quad . \quad . \quad . \quad . \quad . \qquad (416)$$

Thus the ratio of the wave-length λ to the radius of curvature r is :

$$\frac{\lambda}{r} = \frac{h}{e\sqrt{mr}} \quad . \quad . \quad . \quad . \quad . \quad . \qquad (417)$$

If this ratio is to assume an appreciable value the numerator and the denominator of the fraction must be of the same order of magnitude, that is :

$$h \sim e\sqrt{mr} \quad . \quad . \quad . \quad . \quad . \quad . \qquad (418)$$

If, in accordance with our above assumption, we set r equal to the value 10^{-7} cm., then in electrostatic units:

$$e = 4{\cdot}77 \,.\, 10^{-10}(\text{erg} \,.\, \text{cm})^{\frac{1}{2}}, \quad m = 9{\cdot}02 \,.\, 10^{-28}\text{grm.}$$
$$h \sim 4{\cdot}5 \,.\, 10^{-27}\text{erg. sec.} \quad . \quad . \quad . \quad . \qquad (419)$$

The more exact value of the quantum of action is :

$$h = 6{\cdot}55 \,.\, 10^{-27}\text{erg. sec.} \quad . \quad . \quad . \quad . \qquad (420)$$

INDEX

PRINTED IN GREAT BRITAIN BY RICHARD CLAY & SONS, LIMITED,
BUNGAY, SUFFOLK.

b